

**UNIVERSIDADE FEDERAL DO PARANÁ**

**EDUARDO DIMAS ANDRINO NOGUEIRA**

**ANÁLISE DE *BRAND EQUITY* SOB A PERSPECTIVA DO CONSUMIDOR NAS  
MÍDIAS SOCIAIS POR MEIO DA MINERAÇÃO DE OPINIÃO E ANÁLISE DE  
REDES SOCIAIS**

**CURITIBA  
2015**

**EDUARDO DIMAS ANDRINO NOGUEIRA**

**ANÁLISE DE *BRAND EQUITY* SOB A PERSPECTIVA DO CONSUMIDOR NAS  
MÍDIAS SOCIAIS POR MEIO DA MINERAÇÃO DE OPINIÃO E ANÁLISE DE  
REDES SOCIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação, Área de concentração: Gestão da Informação e do Conhecimento, do Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Ciência, Gestão e Tecnologia da Informação.

Orientadora: Prof<sup>a</sup>. Dra. Denise Fukumi Tsunoda

**CURITIBA  
2015**

UNIVERSIDADE FEDERAL DO PARANÁ. SISTEMA DE BIBLIOTECAS.  
CATALOGAÇÃO NA FONTE

Nogueira, Eduardo Dimas Andrino, 1986-  
Análise de *brand equity* sob a perspectiva do consumidor nas mídias  
sociais por meio da mineração de opinião e análise de redes sociais /  
Eduardo Dimas Andrino Nogueira. - 2015.

235 f.

Orientadora: Denise Fukumi Tsunoda.

Dissertação (Mestrado) - Universidade Federal do Paraná.  
Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da  
Informação, do Setor de Ciências Sociais Aplicadas.  
Defesa: Curitiba, 2015.

1. Comportamento do consumidor. 2. Mídia social. 3. Redes sociais  
on-line. 4. Marca de produtos - Marketing. I. Tsunoda, Denise Fukumi,  
1972-. II. Universidade Federal do Paraná. Setor de Ciências Sociais  
Aplicadas. Programa de Pós-Graduação em Ciência, Gestão e  
Tecnologia da Informação. III. Título.

CDD 658.827



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO PARANÁ  
Setor CIÊNCIAS SOCIAIS APLICADAS  
Programa de Pós Graduação em CIÊNCIA, GESTÃO E TECNOLOGIA DA  
INFORMAÇÃO

Código CAPES: 40001016058P1

### PARECER DA BANCA EXAMINADORA

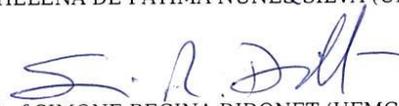
Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **EDUARDO DIMAS ANDRINO NOGUEIRA**, intitulada: "**ANÁLISE DE BRAND EQUITY SOB A PERSPECTIVA DO CONSUMIDOR NAS MÍDIAS SOCIAIS POR MEIO DA MINERAÇÃO DE OPINIÃO E ANÁLISE DE REDES SOCIAIS**", após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO, completando-se assim todos os requisitos previstos nas normas desta Instituição para a obtenção do Grau de **Mestre em CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO**.

CURITIBA, 10 de Dezembro de 2015.

  
Prof DENISE FUKUMI TSUNODA (UFPR)  
(Presidente da Banca Examinadora)

  
Prof DEBORAH RIBEIRO CARVALHO (UFPR)

  
Prof HELENA DE FÁTIMA NUNES SILVA (UFPR)

  
Prof SIMONE REGINA DIDONET (UFMG)

*Dedico esse trabalho a minha esposa,  
Pâmela, pela compreensão e apoio  
incondicionais prestados durante o  
desenvolvimento dessa pesquisa e aos  
meus pais, Mario e Izabel, pelo esforço e  
dedicação que me permitiram desfrutar  
desse momento.*

## **AGRADECIMENTOS**

À Deus, por iluminar meu caminho e me abençoar com saúde, paz e inspiração, tornando essa estrada mais amena e me ajudando a enfrentar os desafios encontrados.

À Professora Dra. Denise Fukumi Tsunoda, pelas orientações, conselhos e recomendações prestados durante o desenvolvimento dessa pesquisa, mantendo sempre acesa em mim a chama da busca pelo conhecimento.

À Professora Dra. Helena de Fátima Nunes Silva por ser a incubadora dessa pesquisa ainda em sua fase embrionária, possibilitando que ela se desenvolvesse de forma sólida e consistente.

Às Professoras Dra. Deborah Ribeiro Carvalho e Dra. Simone Regina Didonet, pelo compartilhamento de seus conhecimentos, contribuindo sobremaneira para a construção desse trabalho.

Aos colegas do Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação, pelas discussões que foram fundamentais para o nascimento dessa pesquisa e também pelos cafés nos intervalos entre as aulas.

Aos colegas de trabalho da Divisão de Sistemas de Informação da Universidade Federal do Paraná, pelo apoio e incentivo prestados, em especial a Antônio Rodrigues Barros, pela compreensão e paciência durante esses dois anos de Mestrado.

Aos familiares e amigos que sempre estiveram comigo, mesmo que por vezes fisicamente distantes.

A todos que de alguma forma contribuíram para o desenvolvimento desse trabalho, seja por meio de uma rápida conversa informal ou pelos momentos de descontração, meu muito obrigado!

*One's mind stretched to a new idea never  
goes back to its original dimensions.*

– *Oliver Wendell Holmes, Sr.*

## RESUMO

As redes sociais online expandiram suas “fronteiras virtuais”, tornando a Internet mais um ambiente de interação social do que uma ferramenta de pesquisa. Entretanto, mesmo antes do surgimento e expansão das mídias sociais, profissionais da área de marketing estiveram interessados em identificar a percepção dos consumidores em relação às marcas, o que passou a ser conceituado como equidade da marca baseada no consumidor. Assim, modelos operacionais foram propostos para viabilizar essa tarefa. Esses modelos, contudo, podem ser dispendiosos e de aplicação inconveniente. Buscando solucionar esse problema, esse trabalho teve como objetivo analisar a equidade da marca na perspectiva do consumidor a partir das mídias sociais. Para atingir esse objetivo, foram executadas as seguintes etapas: 1) levantamento das principais dimensões que compõem os modelos conceituais de equidade da marca; 2) proposta de um modelo preliminar para análise da equidade das marcas a partir das mídias sociais; 3) coleta dos dados referentes a quatro marcas de cosméticos presentes na rede social Twitter; 4) preparação dos dados, incluindo a atribuição de polaridade às publicações coletadas; 5) aplicação do modelo preliminar sobre os dados referentes as quatro marcas de cosméticos coletadas; 6) análise dos resultados do modelo preliminar e identificação de pontos de melhoria; 7) proposta de um novo modelo a partir dos pontos de melhoria identificados, contemplando cinco dimensões da equidade da marca; 8) coleta dos dados referentes a quatro marcas de indústrias automobilísticas presentes na rede social Twitter; 9) preparação dos dados, incluindo a atribuição de polaridade às publicações coletadas e categorização dos perfis em duas classes: consumidores e não-consumidores; 10) aplicação do modelo proposto sobre os dados referentes as quatro marcas de indústrias automobilísticas. A aplicação do modelo proposto tornou possível a análise de cinco dimensões da equidade da marca a partir de dados coletados da mídia social Twitter: conhecimento da marca, lealdade à marca, sentimento percebido, qualidade percebida e associações à marca; o procedimento de classificação de perfis em duas categorias (consumidores e não-consumidores) atingiu uma taxa de acerto de 86,5%, utilizando-se o algoritmo C4.5; a classificação de polaridade das publicações, por sua vez, atingiu uma precisão de 81,2%, com emprego da abordagem baseada em aprendizado de máquina, utilizando-se o algoritmo SVM Linear. Os resultados obtidos com a aplicação do modelo proposto demonstraram que a equidade das marcas pode ser analisada a partir de dados obtidos das redes sociais virtuais, levando à descoberta de como os consumidores percebem as marcas nesse ambiente, sem a necessidade de aplicação de questionários, viabilizando que diferentes marcas em diferentes contextos sejam analisadas a partir de critérios objetivos e replicáveis para cada um dos elementos de equidade da marca que compõem o modelo.

Palavras-chave: CBBE; descoberta de conhecimento; análise de sentimentos

## ABSTRACT

*Online social networks expanded their "virtual borders", making the Internet more an environment of social interaction than a research tool. However, even before the emergence and expansion of social media, marketing professionals were interested in identifying the perception of consumers in relation to brands, which came to be conceptualized as consumer-based brand equity. Thus, operational models have been proposed to facilitate this task. These models, however, can be expensive and inconvenient. To overcome this problem, this study aimed to analyze the brand equity from a consumer perspective from social networks through the use of opinion mining techniques and social network analysis. To achieve this goal, the following steps were performed: 1) the main dimensions that make up the conceptual models of brand equity were surveyed; 2) a preliminary model for analysis of the brand equity from the social media has been proposed; 3) data related to four cosmetic brands from the Twitter were collected; 4) the data were prepared, including the assignment of polarity to the collected publications; 5) the preliminary model was applied to the data collected related to the four cosmetic brands 6) the results obtained from the application of the preliminary model were analyzed and areas for improvement were identified; 7) a model was proposed from the identified points of improvement, covering five dimensions of brand equity; 8) data related to four brands of car manufacturers were collected from the Twitter; 9) the data has been prepared, including the assignment of polarity to the collected publications and the categorization of the profiles into two classes: consumers and non-consumers; 10) the proposed model was applied to the data related to four brands of the automotive industries. The application of the proposed model made possible the analysis of five brand equity dimensions from data collected from Twitter: brand awareness, brand loyalty, perceived sentiment, perceived quality and brand associations; the profile classification process into two categories (consumers and non-consumers) had a 86.5% hit rate, using the C4.5 algorithm; polarity classification of publications, in turn, has reached an accuracy of 81.2% with use of approach based on machine learning, using the linear SVM algorithm. The results obtained from the application of the model show that the brand equity can be analyzed from data obtained from virtual social networks, leading to the discovery of how consumers perceive brands in this environment without the need for questionnaires, enabling different brands in different contexts are analyzed by objective and replicable criteria for each of the brand equity elements that make up the model.*

Key-words: *CBBE; knowledge discovery; sentiment analysis*

## LISTA DE FIGURAS

FIGURA 1 – DIMENSÕES DA EQUIDADE DA MARCA SEGUNDO MODELO CONCEITUAL DE KELLER .....	35
FIGURA 2 – DIMENSÕES DA EQUIDADE DA MARCA SEGUNDO MODELO CONCEITUAL DE AAKER E VALORES PROPORCIONADOS PELA EQUIDADE DA MARCA AO CONSUMIDOR E À EMPRESA .....	38
FIGURA 3 – NÍVEIS DE LEALDADE À MARCA SOB A PERSPECTIVA DO CONSUMIDOR .....	39
FIGURA 4 – NÍVEIS DE CONHECIMENTO DA MARCA SOB A PERSPECTIVA DO CONSUMIDOR .....	42
FIGURA 5 – AS QUINZE DIMENSÕES DA QUALIDADE ASSOCIADAS AO PRODUTO E AO SERVIÇO .....	43
FIGURA 6 – ASSOCIAÇÕES AO NOME OU SÍMBOLO DA MARCA CONSTRUÍDAS PELO CONSUMIDOR PARA FORMAÇÃO DA IMAGEM DE UM PRODUTO OU SERVIÇO.....	46
FIGURA 7 – ESQUEMATIZAÇÃO PROPOSTA POR EULLER PARA RESOLUÇÃO DO PROBLEMA DAS SETE PONTES DE KONIGSBERG .....	54
FIGURA 8 – REPRESENTAÇÃO DO PROBLEMA DAS SETE PONTES DE KONIGSBERG ATRAVÉS DE UM GRAFO .....	54
FIGURA 9 – EXEMPLO DE UM GRAFO FORMADO POR UM CONJUNTO DE VÉRTICES E ARESTAS .....	55
FIGURA 10 – REPRESENTAÇÃO FORMAL DE UM GRAFO.....	56
FIGURA 11 – EXEMPLOS DE REPRESENTAÇÃO DE UM (A) MULTIGRAFO E (B) UM PSEUDOGRAFO .....	56
FIGURA 12 – EXEMPLOS DE DÍGRAFOS E GRAFOS ORIENTADOS .....	57
FIGURA 13 – GRÁFOS IDÊNTICOS E ISOMÓRFICOS .....	57
FIGURA 14 – APRESENTAÇÃO DO GRAFO G SEUS SUBGRAFOS G'1, G'2 E G'3 .....	58
FIGURA 15 – OS CONCEITOS DE PASSEIO, TRILHA E CAMINHO EM UM GRAFO .....	59
FIGURA 16 – EXEMPLO DE UM GRAFO DESCONECTADO COM DEZ COMPONENTES .....	60
FIGURA 17 – ALGUNS EXEMPLOS DE ÁRVORES FORMADAS A PARTIR DO GRAFO (A).....	61
FIGURA 18 – MATRIZES COMO FORMA REPRESENTAÇÃO DOS GRAFOS .....	63
FIGURA 19 – VISÃO GERAL DAS ATIVIDADES QUE CONSTITUEM O PROCESSO DE DESCOBERTA DE CONHECIMENTO (KDD) .....	74
FIGURA 20 – A MINERAÇÃO DE DADOS COMO UMA IMPORTANTE ETAPA DO PROCESSÓ KDD E PRINCIPAIS TAREFAS E TÉCNICAS RELACIONADAS .....	76
FIGURA 21 – ÁRVORE DE DECISÃO PARA INFERÊNCIA SOBRE OCORRENCIA DE UM JOGO DE TÊNIS .....	79
FIGURA 22 – SOLUÇÃO BASEADA NO SVM PARA DUAS CLASSES LINEARMENTE SEPARÁVEIS .....	80
FIGURA 23 – CARACTERIZAÇÃO DA PESQUISA QUANTO AO PROPÓSITO, NATUREZA DOS DADOS, FORMA DE OBTEBNÇÃO DOS DADOS, DELINEAMENTO E NATUREZA DA PESQUISA .....	94

FIGURA 24 – PASSOS EXECUTADOS E FERRAMENTAS UTILIZADAS DURANTE AS ETAPAS DE TESTE E PRÉ-TESTE PARA ALCANCE DOS OBJETIVOS PROPOSTOS PARA A PESQUISA .....	102
FIGURA 25 – PASSOS EXECUTADOS NO PRÉ-TESTE DURANTE A ETAPA DE PREPARAÇÃO DOS DADOS .....	106
FIGURA 26 – EXEMPLOS DE TWEETS DUPLICADOS RESULTANTES DE COLETAS SUBJACENTES.....	107
FIGURA 27 – EXECUÇÃO DA FUNÇÃO COUNT AND MERGE DUPLICATE EDGES, EXECUTADA PARA REMOÇÃO DOS TWEETS DUPLICADOS .....	108
FIGURA 28 – EXEMPLOS DE USUÁRIOS DUPLICADOS RESULTANTES DE COLETAS SUBJACENTES.....	109
FIGURA 29 – EXECUÇÃO DA FUNÇÃO REMOVE DUPLICATAS, EXECUTADA PARA REMOÇÃO DOS USUÁRIOS DUPLICADOS.....	110
FIGURA 30 – PASSOS EXECUTADOS DURANTE PROCESSO DE TREINAMENTO SUPERVISIONADO E NÃO-SUPERVISIONADO NA FASE DE PRÉ-TESTE .....	112
FIGURA 31 – PROCESSO DE TREINAMENTO SUPERVISIONADO .....	112
FIGURA 32 – PASSOS EXECUTADOS DURANTE PROCESSO DE TREINAMENTO SUPERVISIONADO E NÃO-SUPERVISIONADO NA FASE DE PRÉ-TESTE .....	115
FIGURA 33 – PASSOS EXECUTADOS NO TESTE DURANTE A ETAPA DE PREPARAÇÃO DOS DADOS .....	119
FIGURA 34 – ÁRVORE GERADA PELO ALGORITMO C4.5 (EM SUA IMPLEMENTAÇÃO J48) PARA CLASSIFICAÇÃO DOS PERFIS DO TWITTER EM DUAS CATEGORIAS: CONSUMIDORES E NÃO-CONSUMIDORES.....	126
FIGURA 35 – ESQUEMATIZAÇÃO DO MODELO PRELIMINAR PARA ANÁLISE DA EQUIDADE DAS MARCAS NAS REDES SOCIAIS, CONTEMPLANDO QUATRO DIMENSÕES: CONHECIMENTO DA MARCA, LEALDADE À MARCA, QUALIDADE PERCEBIDA E ASSOCIAÇÕES À MARCA .	128
FIGURA 36 – PROPOSTA DO MODELO PRELIMINAR PARA ESTABELECIMENTO DO CONHECIMENTO DA MARCA A PARTIR DE UMA REDE DE SEGUIDORES .....	130
FIGURA 38 – PROPOSTA DO MODELO PRELIMINAR PARA ESTABELECIMENTO DA QUALIDADE PERCEBIDA A PARTIR DE UMA REDE DE MENSAGENS .....	135
FIGURA 39 – REDE DE TWEETS RELACIONADOS À MARCA JEQUITI CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	141
FIGURA 40 – REDE DE TWEETS RELACIONADOS À MARCA NATURA CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	144
FIGURA 41 – REDE DE TWEETS RELACIONADOS À MARCA AVON CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	147
FIGURA 42 – REDE DE TWEETS RELACIONADOS À MARCA O BOTICÁRIO CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	149
FIGURA 43 – NUVEM DE PALAVRAS ASSOCIADAS À MARCA AVON CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	152
FIGURA 45 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE AVON CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	154

FIGURA 46 – NUVEM DE PALAVRAS ASSOCIADAS À MARCA O BOTICÁRIO CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	156
FIGURA 48 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE O BOTICÁRIO CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS .....	158
FIGURA 50 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS POSITIVOS SOBRE JEQUITI CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	161
FIGURA 52 – NUVEM DE PALAVRAS ASSOCIADAS À MARCA NATURA CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	164
FIGURA 54 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE NATURA CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	166
FIGURA 55 – ESQUEMATIZAÇÃO DO MODELO PROPOSTO PARA ANÁLISE DA EQUIDADE DAS MARCAS NAS REDES SOCIAIS, CONTEMPLANDO CINCO DIMENSÕES: CONHECIMENTO DA MARCA, LEALDADE À MARCA, QUALIDADE PERCEBIDA, SENTIMENTO PERCEBIDO E ASSOCIAÇÕES À MARCA .....	169
FIGURA 56 – CORRESPONDÊNCIA ENTRE O CONHECIMENTO DA MARCA NORMALIZADO E OS NÍVEIS DE CONHECIMENTO DE AAKER (1998) .....	170
FIGURA 57 – CORRESPONDÊNCIA ENTRE A LEALDADE NORMALIZADA À MARCA E OS NÍVEIS DE LEALDADE DE AAKER (1998) .....	172
FIGURA 59 – REDE DE TWEETS RELACIONADOS À MARCA CHEVROLET CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	182
FIGURA 60 – REDE DE TWEETS RELACIONADOS À MARCA FIAT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS .....	184
FIGURA 62 – REDE DE TWEETS RELACIONADOS À MARCA VOLKSWAGEN CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	187
FIGURA 64 – EXEMPLO DE TWEET RELACIONADO À PALAVRA “HALL” .....	189
FIGURA 65 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS POSITIVOS SOBRE A MARCA CHEVROLET CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS .....	190
FIGURA 66 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE A MARCA CHEVROLET CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS .....	191
FIGURA 68 – NUVEM DE PALAVRAS RELACIONADAS À MARCA FIAT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	193
FIGURA 70 – EXEMPLOS DE TWEET POSITIVO CONTENDO A PALAVRA “STRADA” .....	195
FIGURA 71 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE A MARCA FIAT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS .....	196
FIGURA 73 – EXEMPLOS DE TWEETS NEGATIVOS, ASSOCIADOS A PROCESSOS DE IMPEACHMENT, CONTENDO A PALAVRA “FIAT” .....	198
FIGURA 74 – EXEMPLOS DE TWEETS NEGATIVOS, ASSOCIADOS A ROUBOS OU ACIDENTES, CONTENDO A PALAVRA “UNO” .....	199
FIGURA 75 – NUVEM DE PALAVRAS RELACIONADAS À MARCA RENAULT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	200

FIGURA 77 – EXEMPLOS DE TWEETS POSITIVOS CONTENDO AS PALAVRAS “RENAULT” E “DUSTER” .....	202
FIGURA 78 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE A MARCA RENAULT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS .....	203
FIGURA 80 – EXEMPLOS DE TWEETS NEGATIVOS CONTENDO A PALAVRA “PLACA” .....	205
FIGURA 81 – NUVEM DE PALAVRAS RELACIONADAS À MARCA VOLKSWAGEN CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS.....	206
FIGURA 82 – EXEMPLOS DE TWEETS RELACIONADOS AO ACIDENTE NA FÁBRICA DA VOLKSWAGEN NA ALEMANHA.....	207
FIGURA 84 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE A MARCA VOLKSWAGEN CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS .....	209
FIGURA 86 – EXEMPLOS DE TWEETS PUBLICADOS SOBRE UMA CURIOSIDADE SOBRE A MARCA VOLKSWAGEN .....	211
FIGURA 87 – REDE DE PALAVRAS OBTIDA PELA APLICAÇÃO DO MODELO PROPOSTO PARA O CRITÉRIO QUALIDADE PERCEBIDA, EM QUE SÃO APRESENTADOS OS ADJETIVOS MAIS ASSOCIADOS ÀS MARCAS PESQUISADAS .....	212
FIGURA 88 – RELACIONAMENTO ENTRE AS ETAPAS DA PESQUISA, PASSOS EXECUTADOS, OBJETIVOS ESPECÍFICOS E OBJETIVO GERAL.....	217

## LISTA DE GRÁFICOS

GRÁFICO 1 – PUBLICAÇÕES ENTRE 1989 E 2014 RELACIONADAS AO TEMA PRINCIPAL DA PESQUISA .....	23
GRÁFICO 2 – PRINCIPAIS MOTIVADORES PARA O USO DO TWITTER COMO FERRAMENTA DE MARKETING.....	25

## LISTA DE TABELAS

TABELA 1 – CONSULTAS REALIZADAS ENTRE 04/03/2015 E 04/05/2015 E NÚMERO DE TWEETS E USUÁRIOS RETORNADOS PARA CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS PESQUISADAS.	104
TABELA 2 – NÚMERO DE TWEETS COLETADOS PARA FORMAÇÃO DO CONJUNTO DE TREINAMENTO DA ANÁLISE DE SENTIMENTOS (AGRUPADOS POR POLARIDADE ATRIBUÍDA) .....	104
TABELA 3 – NÚMERO DE SEGUIDORES OBTIDOS A PARTIR DA COLETA DE CADA UM DOS PERFIS ASSOCIADOS ÀS MARCAS DE COSMÉTICOS ANALISADAS.....	105
TABELA 4 – COMPARATIVO ENTRE A QUANTIDADE DE TWEETS ANTES E APÓS A DEDUPLICAÇÃO, REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS .....	108
TABELA 5 – COMPARATIVO ENTRE A QUANTIDADE DE USUÁRIOS ANTES E APÓS A DEDUPLICAÇÃO, REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS .....	110
TABELA 6 – MATRIZ DE CONFUSÃO OBTIDA COMO SAÍDA DO PROCESSO DE TREINAMENTO SUPERVISIONADO COM EMPREGO DO ALGORITMO SVM DURANTE A FASE DE PRÉ-TESTE .....	114
TABELA 7 – MATRIZ DE CONFUSÃO OBTIDA COMO SAÍDA DO PROCESSO DE TREINAMENTO NÃO-SUPERVISIONADO COM EMPREGO DO ALGORITMO SVM DURANTE A FASE DE PRÉ-TESTE .....	114
TABELA 8 – CONSULTAS REALIZADAS ENTRE 06/07/2015 E 03/08/2015 E NÚMERO DE TWEETS E USUÁRIOS RETORNADOS PARA CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS PESQUISADAS .....	118
TABELA 9 – NÚMERO DE SEGUIDORES OBTIDOS A PARTIR DA COLETA DE CADA UM DOS PERFIS ASSOCIADOS ÀS MARCAS AUTOMOBILÍSTICAS ANALISADAS.....	118
TABELA 10 – COMPARATIVO ENTRE A QUANTIDADE DE TWEETS ANTES E APÓS A DEDUPLICAÇÃO, REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS ANALISADAS.....	120
TABELA 11 – COMPARATIVO ENTRE A QUANTIDADE DE USUÁRIOS ANTES E APÓS A DEDUPLICAÇÃO, REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS ANALISADAS.....	121
TABELA 12 – MATRIZ DE CONFUSÃO OBTIDA COMO SAÍDA DO PROCESSO DE TREINAMENTO SUPERVISIONADO COM EMPREGO DO ALGORITMO SVM, DURANTE A FASE DE TESTE .....	122
TABELA 13 – COMPARATIVO ENTRE O NÚMERO DE USUÁRIOS E TWEETS ANTES E APÓS A ELIMINAÇÃO DOS PERFIS CLASSIFICADOS COMO NÃO-CONSUMIDORES, AGRUPADOS PELAS MARCAS AUTOMOBILÍSTICAS ANALISADAS .....	127
TABELA 14 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PRELIMINAR PARA O CRITÉRIO CONHECIMENTO DA MARCA REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS .....	138

TABELA 15 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PRELIMINAR PARA O CRITÉRIO LEALDADE À MARCA REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS .....	139
TABELA 16 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PRELIMINAR PARA O CRITÉRIO QUALIDADE PERCEBIDA REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS .....	140
TABELA 17 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PARA O CRITÉRIO CONHECIMENTO DA MARCA REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS PESQUISADAS .....	177
TABELA 19 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PARA O CRITÉRIO LEALDADE À MARCA REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS PESQUISADAS ...	178
TABELA 20 – CONVERSÃO DOS VALORES OBTIDOS PARA LEALDADE RELATIVA DAS MARCAS PARA OS NÍVEIS DE LEALDADE DA MARCA.....	179
TABELA 21 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PARA O CRITÉRIO SENTIMENTO PERCEBIDO REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS PESQUISADAS .....	180

## LISTA DE QUADROS

QUADRO 1 – PERSPECTIVA HISTÓRICA DOS PRINCIPAIS <i>SITES</i> DE REDES SOCIAIS, APRESENTANDO SEUS PROPÓSITOS, ENDEREÇOS E NÚMERO DE USUÁRIOS.....	70
QUADRO 2 – DESCRIÇÃO DOS ATRIBUTOS RELACIONADOS AOS TWEETS DISPONIBILIZADOS PELA FERRAMENTA NODEXL .....	97
QUADRO 3 – DESCRIÇÃO DOS ATRIBUTOS DISPONIBILIZADOS PELA FERRAMENTA NODEXL REFERENTES AOS USUÁRIOS DA REDE SOCIAL TWITTER .....	98
QUADRO 4 – PROCEDIMENTOS METODOLÓGICOS ADOTADOS DE ACORDO COM OS OBJETIVOS ESPECÍFICOS .....	101
QUADRO 5 – AMOSTRA PARCIAL DA BASE DE USUÁRIOS UTILIZADA NO PROCESSO DE CLASSIFICAÇÃO DOS PERFIS EM DUAS CATEGORIAS: CONSUMIDORES E NÃO-CONSUMIDORES .....	124
QUADRO 6 – EXEMPLOS DE TWEETS DO PERFIL OFICIAL DA MARCA JEQUITI .....	142
QUADRO 7 – EXEMPLOS DE TWEETS DOS PERFIS COM MAIOR PRESTÍGIO NA REDE DA MARCA NATURA.....	145
QUADRO 8 – EXEMPLOS DE TWEETS DOS PERFIS COM MAIOR PRESTÍGIO NA REDE DA MARCA AVON .....	148
QUADRO 9 – EXEMPLOS DE TWEETS DOS PERFIS COM MAIOR PRESTÍGIO NA REDE DA MARCA O BOTICÁRIO.....	150
QUADRO 10 – PALAVRAS MAIS FREQUENTES NOS TWEETS RELACIONADOS À MARCA AVON .....	151
QUADRO 11 – PALAVRAS MAIS FREQUENTES NOS TWEETS RELACIONADOS À MARCA O BOTICÁRIO .....	155
QUADRO 12 – PALAVRAS MAIS FREQUENTES NOS TWEETS RELACIONADOS À MARCA JEQUITI.....	159
QUADRO 13 - PALAVRAS MAIS FREQUENTES NOS TWEETS RELACIONADOS À MARCA NATURA.....	163

## LISTA DE ABREVIATURAS E SIGLAS

AMA	<i>American Marketing Association</i>
API	<i>Application Programming Interface</i>
B2C	<i>Business to Consumer</i>
CBBE	<i>Customer-Based Brand Equity</i>
CSV	<i>Comma Separated Values</i>
DD-CSS	<i>Data-Driven Computational Social Science</i>
FBBE	<i>Firm-Based Brand Equity</i>
KDD	<i>Knowledge Discovery in Databases</i>
SVM	<i>Support Vector Machine</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>19</b>
1.1	PROBLEMA DE PESQUISA	20
1.2	OBJETIVOS	22
1.3	JUSTIFICATIVA	22
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>28</b>
2.1	BRAND EQUITY	28
2.1.1	Marcas: histórico e conceitos	28
2.1.2	O valor da marca: principais conceitos	31
2.1.3	Valor das marcas na perspectiva do consumidor: principais modelos	34
2.2	REDES SOCIAIS	50
2.2.1	Análise de redes sociais	51
2.2.2	Elementos fundamentais	51
2.2.3	Representação por meio de grafos	53
2.2.4	Representação por meio de matrizes	61
2.2.5	Métricas para análise de redes sociais	63
2.2.6	Redes sociais virtuais	69
2.3	MINERAÇÃO DE OPINIÃO	72
2.3.1	Dado, informação e conhecimento	73
2.3.2	Descoberta de conhecimento	74
2.3.3	Mineração de dados	75
2.3.4	Mineração de opinião: definição, tarefas e abordagens	80
2.3.5	Mineração de opinião no Twitter	89
2.4	ALINHAMENTO TEÓRICO	91
<b>3</b>	<b>METODOLOGIA DA PESQUISA</b>	<b>94</b>
3.1	CARACTERIZAÇÃO DA PESQUISA	94
3.2	AMBIENTE DA PESQUISA	95
3.3	FERRAMENTAS	96
3.3.1	NodeXL	96
3.3.2	Rapidminer	99
3.3.3	Weka	99
3.3.4	Tagul	100
3.3.5	DD-CSS	100
3.4	PROCEDIMENTOS METODOLÓGICOS	100
3.4.1	Pré-teste	103
3.4.2	Teste	117
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>128</b>
4.1	MODELO PRELIMINAR	128
4.1.1	Conhecimento da Marca	129
4.1.2	Lealdade à Marca	131
4.1.3	Qualidade Percebida	134
4.1.4	Associações à Marca	137
4.2	RESULTADOS DA APLICAÇÃO DO MODELO PRELIMINAR	138

4.3	CONSIDERAÇÕES SOBRE O MODELO PRELIMINAR.....	167
4.4	MODELO PROPOSTO .....	169
4.4.1	Conhecimento da marca .....	169
4.4.2	Lealdade à marca.....	171
4.4.3	Qualidade percebida .....	172
4.4.4	Sentimento percebido.....	173
4.4.5	Associações à marca .....	176
4.5	RESULTADOS DA APLICAÇÃO DO MODELO PROPOSTO .....	176
4.6	CONSIDERAÇÕES SOBRE O MODELO PROPOSTO .....	213
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>216</b>
5.1	CONFRONTAÇÃO ENTRE OS OBJETIVOS E OS RESULTADOS.....	216
5.2	CONTRIBUIÇÕES DO TRABALHO.....	220
5.3	LIMITAÇÕES DA PESQUISA E SUGESTÕES DE TRABALHOS FUTUROS.....	222
	<b>REFERÊNCIAS.....</b>	<b>224</b>

# 1 INTRODUÇÃO

As redes sociais online como, por exemplo, o Twitter e o Facebook, expandiram suas “fronteiras virtuais” e vem conquistando maior número de usuários. Só no Brasil, segundo dados da agência eMarketeer (2014), eram 78,3 milhões de adeptos às redes sociais em 2013. A previsão da agência é de que em 2017 o número de brasileiros nessas redes alcance a marca de 110 milhões, o que deverá representar mais da metade da população do Brasil projetada pelo IBGE (2008) para aquele ano.

Confirmando esse rápido crescimento, a rede social Facebook ultrapassou o site Google em número de acessos no Brasil pela primeira vez em abril de 2012 (DIAS, 2012): a internet passava a se tornar, a partir de então, mais um ambiente de interação social do que uma ferramenta de pesquisa. Tal interação, por meio do compartilhamento e troca de mensagens, é responsável pela produção de um enorme volume de dados na web: são realizadas, em média, 500 milhões de postagens no Twitter por dia (TWITTER, 2013). A rede social Facebook, por sua vez, armazena mais de 300 petabytes em dados (TRAVERSO, 2013) referentes ao perfil de seus usuários, postagens e compartilhamentos.

Esses números enfatizam a presença das redes sociais no dia a dia das pessoas, fato que tem motivado profissionais das mais diversas áreas a voltarem seus olhares sobre essa grande fatia da web. Saber sobre o que as pessoas estão falando, como e com quem elas interagem ou qual a opinião destas sobre determinado assunto, têm levado gestores, consultores de marketing, publicitários, sociólogos, cientistas e profissionais de tecnologia da informação a se debruçar sobre as redes sociais com o objetivo de descobrir as informações em potencial escondidas pelo grande volume de dados e interações entre seus usuários.

Em vista disso, essa pesquisa pretende compreender as mídias sociais e seus usuários, com um viés notadamente mercadológico e, a partir disso, extrair conhecimento útil e relevante que possa ser utilizado pelos profissionais que atuam nessa área. Nesse sentido, foi desenvolvido o problema de pesquisa, apresentado na seção 1.1.

## 1.1 PROBLEMA DE PESQUISA

Mesmo antes do surgimento e expansão das mídias sociais, profissionais da área de marketing estiveram interessados em identificar a percepção dos consumidores em relação às marcas. No início dos anos 90, Keller (1993) apresentou o conceito de equidade da marca sob a perspectiva do cliente, também conhecida como *Customer-Based Brand Equity* (CBBE), que foi definido pelo autor como o “efeito diferencial do conhecimento da marca na resposta do consumidor ao marketing de uma marca” (KELLER, 1993, p. 8). Esse conhecimento é baseado em dois componentes: lembrança e imagem da marca.

Mais tarde, Aaker (1998) complementou o modelo anterior ao propor um paradigma conceitual de *brand equity*, adicionando mais duas dimensões – lealdade e qualidade percebida – a esse construto.

Embora modelos para avaliação da equidade da marca a partir da percepção do consumidor tenham sido propostos por Keller (1993) e Aaker (1998), eles nunca operacionalizaram uma escala para sua medida, fazendo com que uma série de metodologias fossem propostas no meio acadêmico para quantificação desse constructo. A maior parte dessas metodologias emprega complexos procedimentos estatísticos, tornando-as de difícil compreensão pelos profissionais de marketing (CHRISTODOULIDES e CHERNATONY, 2010).

Esses experimentos empíricos para operacionalização da equidade de marca sob a perspectiva do consumidor podem ser baseados em abordagens diretas ou indiretas de medição (CHRISTODOULIDES e CHERNATONY, 2010). Os modelos operacionais de abordagem direta buscam medir a equidade da marca focando diretamente nas preferências dos consumidores (KAMAKURA e RUSSEL, 1993; SWAIT et al, 1993; PARK e SRINIVASAN, 1994; SHANKAR et al, 2008) ao passo que abordagens indiretas medem a equidade da marca a partir de suas manifestações demonstráveis (LASSAR, 1995; YOO e DONTU, 2001; VAZQUEZ, DEL RIO e IGLESIAS, 2002; PAPPU, QUESTER e COOKSEY, 2005).

Com esse objetivo, Yoo e Donthu (2001) selecionaram doze marcas referentes a três categorias distintas de produtos (filmes para câmeras, tênis esportivos e aparelhos de TV) e aplicaram questionários utilizando uma amostra composta por 460 estudantes universitários; para medição da de cada um dos elementos relacionados à equidade das marcas, cada questionário continha 48 questões, cujas respostas

variavam conforme escala Likert de cinco pontos, onde o menor nível correspondia à opção “discordo fortemente” e o maior correspondia à opção “concordo fortemente”.

De modo análogo, Pappu, Quester e Cooksey (2005) aplicaram um questionário para consumidores reais (não-estudantes). O questionário era composto por 13 questões relacionadas à equidade da marca, sendo que a confiança na marca foi medida através de questões que utilizavam escala dicotômica (sim/não), enquanto que os demais elementos da equidade da marca foram mensurados por meio de escala de 11 pontos, onde o menor nível representava a opção “discordo fortemente” e o maior, “concordo fortemente”

Aaker (1998), entretanto, salienta que as medidas dos padrões de comportamento do consumidor – como as propostas por Yoo e Donthu (2001) e Pappu, Quester e Cooksey (2005) – podem ter limitações, ao passo que a obtenção das mesmas pode ser inconveniente e dispendiosa, exatamente pelo fato de os dados serem geralmente obtidos por meio da aplicação de questionários.

Desde o advento da internet, profissionais ligados ao marketing buscaram monitorar a rede para extrair informações relacionadas às marcas, seja com o objetivo de se relacionar com seus consumidores ou melhorar sua inteligência competitiva. Entretanto, a fragmentação da mídia e, conseqüentemente, a mudança de comportamento do consumidor, tornaram ultrapassados os métodos tradicionais de monitoramento (KIM, ANDERSON e JOSEPH, 2006).

Essa mudança de paradigma – resultado da grande expansão das redes sociais e do volume de dados gerado pelas interações entre os usuários – tem levado tanto profissionais como empresas da área de marketing a utilizar o conteúdo dessas mídias para tomada de decisão.

Assim, consumidores não estão mais restritos à opinião de um amigo ou conhecido a respeito de um produto na medida em que há uma grande quantidade de informação publicamente disponível (LIU, 2012). Pelo mesmo motivo, as empresas não precisam mais realizar pesquisas de opinião, uma vez que a aplicação de questionários pode ter se tornado obsoleta à medida em que informações podem ser recuperadas automaticamente a partir da web (BOIY e MOENS, 2009; LIU, 2012).

Vale salientar, porém, que esse volume de dados disponível nas redes sociais nem sempre implica na extração de informação verdadeiramente relevante e útil que possa ser utilizada como insumo para tomada de decisão, seja por consumidores, gestores ou demais profissionais interessados nas redes sociais.

Nesse sentido, monitorar as chamadas mídias sociais e, a partir delas, extrair informação relevante, tornou-se um grande desafio, uma vez que analisar os textos manualmente é uma tarefa árdua. Além do mais, ao contrário das bases tradicionais, os registros encontrados na web estão na forma de textos não estruturados (LIU, 2012).

A análise do cenário exposto até aqui, caracterizado, por um lado, pelo crescimento das redes sociais – com suas características intrínsecas e seu grande potencial para descoberta de conhecimento – e por outro, pelo interesse dos profissionais de marketing em medir o valor de uma marca sob a perspectiva dos consumidores, leva-nos à questão de pesquisa: **“Como analisar o *brand equity* na perspectiva do consumidor a partir das mídias sociais?”**.

## 1.2 OBJETIVOS

A fim de responder a questão levantada nessa pesquisa, definiu-se o objetivo geral e, de modo que este possa ser atingido, deverão ser alcançados os objetivos específicos. Assim, este trabalho tem como objetivo geral **analisar o *brand equity* na perspectiva do consumidor a partir das mídias sociais**. Para um melhor detalhamento e entendimento do objetivo geral, definiram-se os seguintes objetivos específicos:

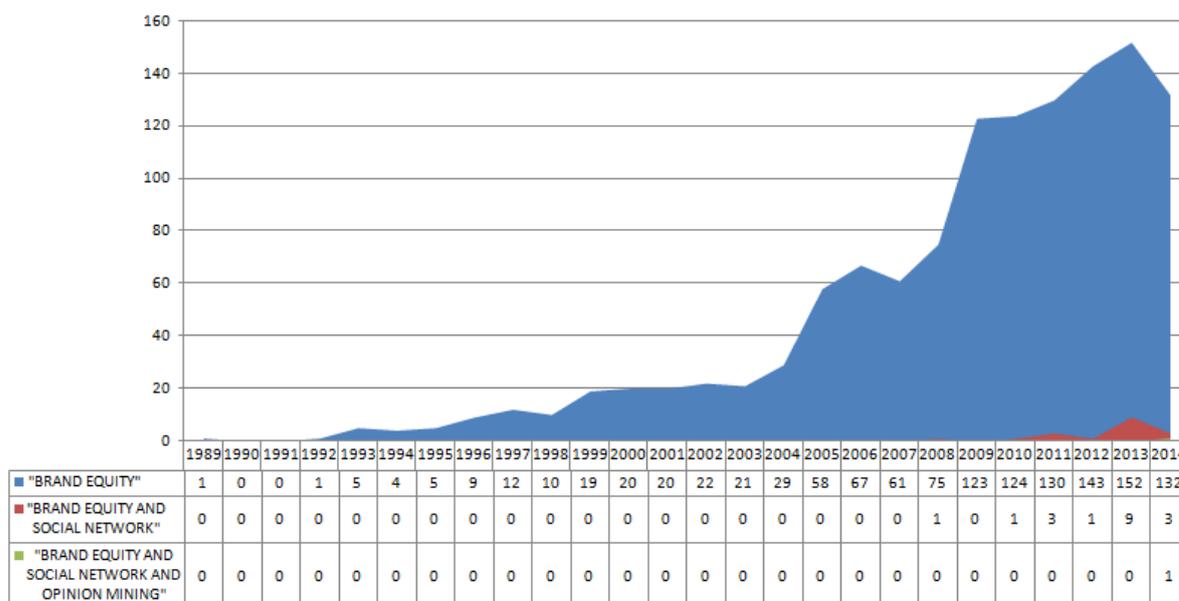
- a) identificar as dimensões presentes nos modelos conceituais de *brand equity* sob a perspectiva do consumidor;
- b) propor um modelo para análise do *brand equity* sob a perspectiva do consumidor a partir das mídias sociais, contemplando as dimensões identificadas nos modelos conceituais;
- c) validar o modelo proposto por meio de aplicação sobre dados gerados por usuários da rede social Twitter.

## 1.3 JUSTIFICATIVA

Em um levantamento no portfólio construído a partir do indexador *Scopus* por meio de pesquisa realizada pela busca do termo “*brand equity*” nos tópicos dos artigos (título, resumo e palavras-chave), foram obtidos 1.242 documentos relacionados a essa temática, publicados no período 1989 a 2014. Entretanto, quando o termo “*brand*

*equity*” foi combinado com “*social network*”, foram recuperados apenas 18 documentos; quando o termo “*opinion mining*” foi combinado com os outros dois, obteve-se apenas um trabalho. A distribuição dos documentos retornados em cada uma das três consultas, por ano de publicação, é apresentada no Gráfico 1.

**GRÁFICO 1 – PUBLICAÇÕES ENTRE 1989 E 2014 RELACIONADAS AO TEMA PRINCIPAL DA PESQUISA**



**FONTE: O AUTOR (2015)**

Percebe-se, a partir do Gráfico 1, que apesar do número de publicações relacionadas ao termo *brand equity* ter apresentado uma tendência de crescimento nos últimos anos – principalmente a partir de 2005, atingindo seu ápice em 2013, com 152 publicações naquele ano – há uma falta de iniciativas no âmbito acadêmico para avaliar a equidade da marca por meio dos dados obtidos a partir redes sociais, o que é ainda mais claro quando a mineração de opinião é incluída nessa análise.

O único documento recuperado que abrange as três temáticas pesquisadas é um artigo intitulado “*Plateforme d’étiquetage dynamique pour la fouille d’opinion*”, cujo objetivo foi analisar como grupos sociais percebem uma determinada marca; para isso, os autores propõem uma plataforma para rotulagem das marcas a partir dos dados sociais da web, baseada na participação colaborativa e mecanismos de aprendizagem estatísticos (REUTENAUER e URVOY, 2014).

Mesmo se desconsiderada a aderência temática relacionada ao uso de técnicas sociométricas – sobretudo a análise de redes sociais – dos 18 trabalhos que

procuram analisar a equidade da marca a partir de dados das mídias sociais, a maior parte utiliza dados de contato dos usuários dessas redes para envio de questionários ou *surveys*, como Barreda (2014), que realiza o envio de questionários a usuários de sites e redes sociais de viagens para estudar a influência da interatividade dessas ferramentas na consciência e imagem das marcas relacionadas a esses domínios.

No que se refere à análise da equidade das marcas a partir de conteúdo gerado pelos usuários dessas mídias, encontrou-se alguma aderência apenas com o trabalho desenvolvido por Nam e Kannan (2014). Nessa pesquisa, os autores procuram investigar a influência de elementos da equidade da marca baseados no consumidor – conhecimento da marca e imagem da marca – sobre o valor financeiro da marca. Como parte desse trabalho, os pesquisadores analisam como rótulos coletados a partir da mídia social *Delicious* podem atuar como intermediários na obtenção do desses elementos-chave.

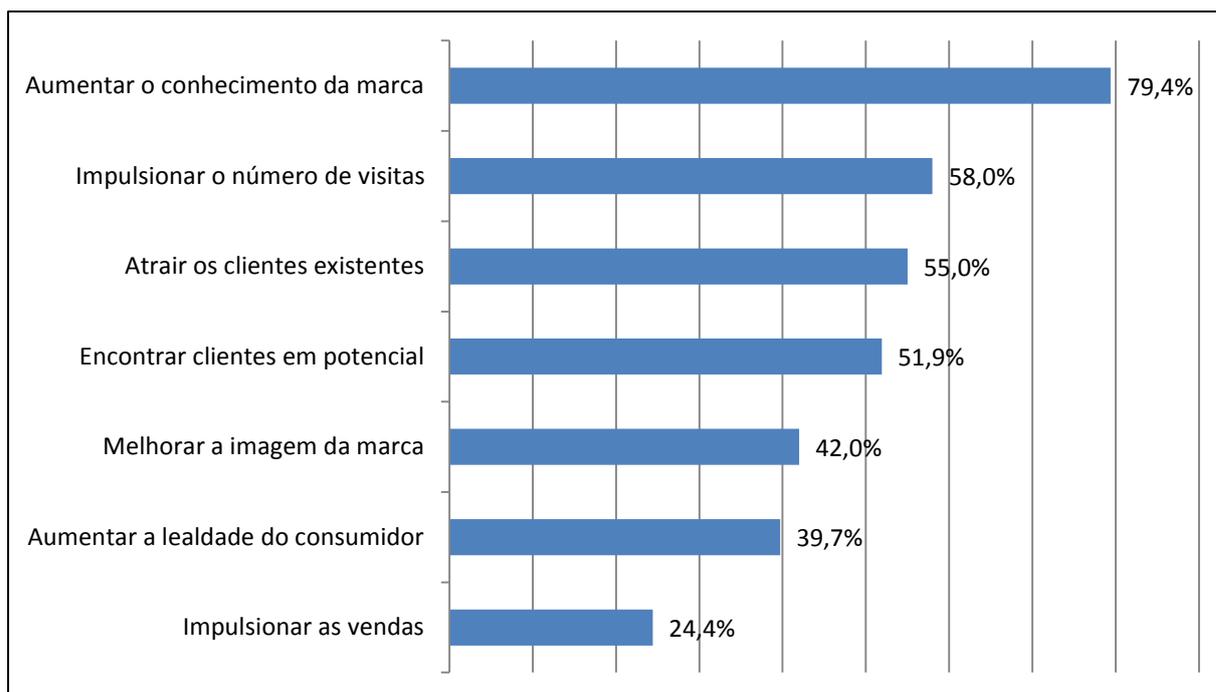
Resultados semelhantes aos observados na base do indexador *Scopus* são encontrados quando a mesma pesquisa é executada na Base de Dados de Teses e Dissertações. Quando o termo “*brand equity*” é buscado, são retornadas 31 dissertações e 4 teses – não havendo nenhum retorno quando o termo é traduzido para português brasileiro. Da mesma forma, a busca não retorna nenhum resultado quando o termo “*brand equity*” é combinado com “*social network*” ou “rede social”, ou ainda com “*opinion mining*” ou “mineração de opinião”.

Portanto, avaliar o *brand equity* sob a perspectiva do consumidor a partir das redes sociais é relevante para a sociedade acadêmica à medida que explora um nicho de pesquisa pouco examinado, caracterizado pela interdisciplinaridade entre as áreas de Negócio, Gestão e Contabilidade com Ciência da Computação e Ciências Sociais.

Apesar do número reduzido de estudos acadêmicos relacionados ao tema, as mídias sociais, sobretudo os microblogs, têm sido amplamente exploradas pelas companhias como parte de sua estratégia de marketing, dado seu potencial como fonte de inteligência competitiva, uma vez que os sentimentos dos consumidores em relação à marca podem ser obtidos quase que em tempo real, além de constituir uma plataforma na qual as marcas podem construir relações com seus clientes (JANSEN et. al., 2009). Esses motivos podem explicar os dados obtidos com pesquisa realizada em 2014 pela organização *Social Media Examiner*, na qual 81% dos 2400 profissionais de marketing entrevistados relataram utilizar o Twitter como parte de seu planejamento de marketing B2C (*Business to Consumer*) (RICHTER, 2014a).

Outra pesquisa realizada no mesmo ano pela *Social Media Marketing University* (RICHTER, 2014b) mostra as principais motivações que levam os profissionais da área mercadológica a utilizar o Twitter como uma ferramenta de marketing. As principais respostas são apresentadas no Gráfico 2.

**GRÁFICO 2 – PRINCIPAIS MOTIVADORES PARA O USO DO TWITTER COMO FERRAMENTA DE MARKETING**



**FONTE: ADAPTADO DE RICHTER (2014B)**

Destacam-se, a partir do Gráfico 2, três motivações: aumentar o conhecimento da marca (79,4%), melhorar a imagem da marca (42%) e aumentar a lealdade do consumidor (39,7%); todas elas superando a motivação por vendas (24,4%). Esses dados demonstram que o Twitter é visto pelos profissionais de marketing como uma importante plataforma para alavancar a equidade da marca por parte do consumidor.

Assim, a proposta de um modelo para um melhor entendimento de como os consumidores percebem uma marca nas redes sociais pode auxiliar os profissionais de marketing a compreender o comportamento desses indivíduos nas novas mídias, por meio da extração de informação útil a partir dos dados produzidos pelos usuários dessas redes e, a partir da análise e interpretação, transformar essa informação em conhecimento que servirá de base para a tomada de decisão por parte desses interessados, melhorando assim seu relacionamento com o cliente e aprimorando sua inteligência competitiva.

Este trabalho justifica-se ainda pelo fato de ser aderente ao Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da informação, à medida que está alinhado tanto com a área de concentração – Gestão da Informação e do Conhecimento – quanto com a linha do Programa – Informação, Tecnologia e Gestão – e pode gerar subsídios para coleta, tratamento e análise de dados provenientes das redes sociais, a partir do desenvolvimento de um método para avaliação da equidade da marca pelos usuários dessas redes. Os resultados alcançados a partir da aplicação do método, técnicas e ferramentas propostas podem ser utilizados pelas organizações para obtenção de vantagem competitiva por meio da melhoria de seus processos, produtos e serviços.

Dentro do Programa, apesar de nenhum trabalho investigar as marcas e seu valor sob a perspectiva do cliente, cinco trabalhos abordaram a temática “redes sociais”.

Souza (2010) propôs a utilização estratégica das redes sociais para amplificar a criação do conhecimento de modo a facilitar o exercício da inovação, indicando “um conjunto de recomendações para as organizações que buscam ser inovadoras e planejam utilizar as redes sociais para amplificar a criação do conhecimento” (SOUZA, 2010, p. 147).

Lima (2011) identificou que os profissionais de comunicação eram otimistas em relação ao uso potencial das redes sociais em seus negócios, embasados, sobretudo, na inovação proporcionada por essas ferramentas e na boa aceitação dessas ferramentas pelo público-alvo, destacando a promoção do relacionamento com públicos distintos como um dos principais benefícios.

Ritzmann (2012) analisou o modelo de uso da informação a partir das redes sociais online como fonte de informação para a criação de significado e concluiu que esse modelo é um mecanismo válido para o objetivo estudado.

Alcântara (2013) explorou técnicas de recuperação e classificação de informações, tanto em sites na web quanto nas redes sociais, desenvolvendo *scripts* para a extração de conteúdo dos sites e utilizando métricas de caracterização de nós em grafos para a classificação das informações resultantes das interações sociais na web.

Santos (2014) analisou os dados automaticamente extraídos de *fan pages* da rede social Facebook por meio do uso de métricas de centralidade para medir a difusão da informação nas redes sociais, sobretudo no contexto do *marketing* de

relacionamento, além de representar graficamente a interatividade entre as marcas relacionadas às páginas e seus clientes.

O presente trabalho pretende dar continuidade a esse tema dentro do Programa, com um viés voltado ao marketing e à luz da descoberta de conhecimento em bases de dados.

Para isso, esse documento é organizado conforme se apresenta a seguir. O tripé teórico dessa pesquisa é apresentado no Capítulo 2, apoiando-se nos temas *brand equity*, redes sociais e mineração de opinião, enquanto que no Capítulo 3 apresenta-se a metodologia de pesquisa, por meio de sua caracterização, definição do ambiente, procedimentos metodológicos adotados, além da descrição dos métodos aplicados nas etapas de coleta e pré-processamento de dados. No Capítulo 4 é apresentado o modelo proposto nessa pesquisa e os resultados obtidos com sua aplicação. Finalmente, no Capítulo 5, são realizadas as considerações finais da pesquisa, apresentando a confrontação entre os objetivos e os resultados, as principais contribuições do trabalho e as sugestões para trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

Nessa seção são apresentados os elementos teóricos que fundamentam essa pesquisa: *brand equity*, redes sociais e mineração de opinião.

### 2.1 BRAND EQUITY

Embora não haja um consenso para definição do seu significado (MARTINS, 2000; KELLER e MACHADO, 2006), são apresentados nesse tópico os principais fundamentos relacionados ao tema *brand equity*. Apesar de ser traduzido algumas vezes para “valor da marca” e “patrimônio da marca” (KELLER e MACHADO, 2006), talvez seja mais bem representado por “equidade da marca”, uma vez que o *brand equity* é apenas um dos termos que constituem o valor de uma marca (*brand value*) (RAGGIO e LEONE, 2007; GRANNELL, 2009).

Entretanto, antes de se apresentar a definição do termo *brand equity* (ou as definições), é preciso caracterizar o que é uma “marca” no contexto dessa pesquisa, conceito que é investigado na seção 2.1.1.

#### 2.1.1 Marcas: histórico e conceitos

Quando o consumidor se depara com a situação de escolher um produto em detrimento de outro – seja quando faz as compras no supermercado, quando vai a uma concessionária de veículos ou quando adquire produtos em um shopping – um dos fatores que exercem influência sobre sua decisão é a marca desses produtos (KELLER, 1993; AAKER, 1998; KELLER e MACHADO, 2006; KAPFERER, 2012).

A marca surgiu justamente dessa necessidade de diferenciar os produtos que eram fabricados, sendo que alguns autores defendem que sua utilização tenha ocorrido pela primeira vez para demarcação de gado e produtos de cerâmica, por volta de 1300 a.C., pela civilização do Antigo Egito (AAKER, 1996; TAVARES, 1998; ACCIOLY, 2000; KELLER e MACHADO, 2006). Após o domínio dos romanos e o crescimento do seu império, esses também passaram a utilizar as marcas para diferenciação de suas mercadorias que eram negociadas com os outros povos, mas foi apenas no renascimento que surgiram os primeiros livros de registros de marcas.

Com a Revolução Industrial, os produtos passaram a ser fabricados em larga escala e marcados de maneira mais uniforme, também com o objetivo de identificar sua origem (AAKER, 1998; SOARES, 1988; ACCIOLY, 2000; DE FREITAS, 2012).

Formalmente, a *American Marketing Association* (AMA) define marca como “um nome, termo, desenho, símbolo ou qualquer outra característica que identifica um vendedor de bens ou serviços como distinto de outros vendedores” (AMA, 2015). Aaker (1998) destaca mais uma vez que o objetivo da marca é diferenciar produtores, deixando claro aos consumidores a origem dos produtos:

Uma marca é um nome diferenciado e/ou um símbolo (tal como um logotipo, marca registrada ou desenho de embalagem) destinado a identificar os bens ou serviços de um vendedor ou de um grupo de vendedores e a diferenciar esses bens e serviços daqueles dos concorrentes. Assim, uma marca sinaliza ao consumidor a origem do produto e protege, tanto o consumidor quanto o fabricante, dos concorrentes que ofereçam produtos que pareçam idênticos. (AAKER, 1998, p. 7)

É preciso, porém, estabelecer a diferença entre uma marca e um produto. Armstrong e Kotler (2004) definem o que é um produto, extrapolando o conceito mais intuitivo de bem material que pode ser comercializado:

Definimos um produto como algo que pode ser oferecido a um mercado para apreciação, aquisição, uso ou consumo e que pode satisfazer um desejo ou necessidade. Produtos incluem mais do que apenas bens tangíveis. Definimos amplamente, incluem objetos físicos, serviços, eventos, pessoas, lugares, organizações, ideias ou um misto de todas essas entidades. (ARMSTRONG e KOTLER, 2004, p. 200)

De acordo com o conceito anterior, uma cidade, um político ou uma ONG podem ser considerados produtos. Entretanto, quando temos uma cidade que oferece diferenciais em relação a outra cidade como, por exemplo, uma melhor qualidade de vida, maior IDH ou um sistema de transporte público mais eficiente, pode-se dizer que há uma marca associada a esse produto que, nesse caso, se trata de uma cidade. Nesse sentido, Keller e Machado (2006) definem o que é uma marca, estabelecendo uma estreita relação com produto:

Uma marca é, portanto, um produto, mas um produto que acrescenta outras dimensões que o diferenciam de algum modo de outros produtos desenvolvidos para satisfazer a mesma necessidade. Essas diferenças podem ser racionais e tangíveis – relacionadas com o desempenho de produto da marca – ou mais simbólicas, emocionais e intangíveis –

relacionadas com o que a marca representa. (KELLER e MACHADO, 2006, p. 4)

Importante destacar que, para os autores, as marcas podem ser diferenciadas tanto por aspectos concretos e palpáveis, mas também por aspectos subjetivos e relacionados ao subconsciente de cada indivíduo. Um consumidor pode, por exemplo, escolher um produto de uma determinada marca porque aquele produto o remete a boas recordações de infância ou experiências agradáveis vivenciadas em seu passado.

Aprofundando esse entendimento, Ambler e Styles (1997, p. 222, tradução nossa) definem marca, adicionando um novo aspecto: o compromisso gerado com o consumidor. Para os autores, uma marca é “a promessa de um pacote de atributos que alguém compra e que proporciona uma satisfação [...] Os atributos que compõem uma marca podem ser reais ou ilusórios, racionais ou emocionais, tangíveis ou invisíveis.”.

Kotler (2000) propõe uma ideia semelhante, também dando destaque para o quase “juramento” proposto pela marca ao consumidor. Segundo o autor, uma marca é uma promessa do fabricante daquele produto em prover um conjunto de qualidades, vantagens e serviços uniformes para seus consumidores.

Logo, uma marca se compromete com o consumidor em entregar algo agradável ao passo que este, por sua vez, gera uma expectativa ao adquirir um produto daquela marca, o que compactua com a interpretação proposta por Schultz e Barnes (2001, p. 44). Para eles, “a marca é mais do que um nome, símbolo ou ícone – é um vínculo que só o consumidor é capaz de criar”. Nesse mesmo viés, Keller (1993) infere que a assimilação sobre a marca que reside no subconsciente do consumidor é o que forma a imagem de uma marca.

Kapferer (2012, p.8) sintetiza: “marca é um nome com poder de influência”. Essa influência das marcas ficou mais explícita nos anos 80 e 90, período em que ocorreu uma onda de fusões e aquisições de diversas empresas, sobretudo nos Estados Unidos. Esse fenômeno “contribuiu para a divulgação, entendimento e medição das marcas, como aquilo que se convencionou chamar de *brand equity*” (MARTINS, 2000, p. 308). As empresas passavam a ser avaliadas, a partir de então, não apenas por seu patrimônio tangível, mas também pelo valor de suas marcas, que passou a ser aceito como um ativo de marketing (SRIVASTAVA, SHERVANI e FAHEY, 1998; RUST, LEMON e ZEITHAML, 2004). Além de um ativo estratégico, a

marca passava a ser também a fonte primária de vantagem competitiva para uma organização (AAKER, 1998).

### 2.1.2 O valor da marca: principais conceitos

Na literatura não há um entendimento comum a respeito do significado do termo *brand equity* (KELLER e MACHADO, 2006; VILLANUEVA e HANSSENS, 2007; CHRISTODOULIDES e DE CHERNATONY, 2010), implicando na falta de consenso sobre como medi-lo (KAPFERER, 2012). Em 1994 havia 26 modelos de medição do *brand equity*; em 2011 eram 300 formas de mensuração (PIRES, 2013). Há uma linha de pesquisadores que o analisam por uma ótica financeira, com o objetivo de comensurar o valor monetário de uma marca. Para outros, entretanto, *brand equity* diz respeito de como o consumidor percebe uma marca e de como isso o influencia no momento da compra. Outros ainda propõem um constructo que concilia ambas as abordagens (CHAVES, 2010; OLIVEIRA, 2013).

Para tentar organizar as múltiplas visões sobre *brand equity*, Christodoulides e De Chernatony (2010) propuseram a existência de duas perspectivas: o *customer-based brand equity* (CBBE) e o *firm-based brand equity* (FBBE).

Alguns autores, como Farquhar (1989), Simon e Sullivan (1993) e Haigh (1999) apresentam uma abordagem mais voltada para a área financeira, ao discutir o valor monetário que uma marca cria para o negócio (CHRISTODOULIDES e DE CHERNATONY, 2010).

Nesse sentido, Farquhar (1989 *apud* OLIVEIRA, 2009, p. 26) afirma que um bom gerenciamento de *brand equity* proporciona “adição de valor para a empresa, para o negócio ou consumidor, na medida em que uma dada marca complementa um produto.”

Simon e Sullivan (1993) conceituam *brand equity* à luz do mercado financeiro, reparando-o como um diferencial de preço das ações concedido à empresa pelos investidores da bolsa de valores que vai além dos ativos tangíveis da marca, ou seja, depois de subtrair o valor patrimonial tangível da capitalização do mercado da empresa, o excesso de capital é o valor de ativos intangíveis da marca, ou seja, seu *brand equity*.

Haigh (1999), analisando através desse mesmo prisma, considera apenas o valor patrimonial de uma marca e destaca que a avaliação do valor de uma marca deve ser apoiada em dados concretos, fazendo uma clara contraposição às metodologias de medição do *brand equity* baseada no consumidor, que leva em consideração as percepções – muitas vezes abstratas e difíceis de ser comensuradas – que este possui sobre uma marca:

Avaliar o valor de uma marca terá crédito apenas se forem baseadas em previsões confiáveis e previsões confiáveis devem ser baseadas em dados estatisticamente e historicamente válidos [...] Quando previsões são apoiadas em evidências robustas, utilizando, por exemplo, dados de modelagem econométrica ou análise de correlação, as avaliações tornam-se um aditivo com credibilidade para a gestão da tomada de decisão. (HAIGH, 1999, p. 30, tradução nossa)

Para outros autores como Leuthesser (1988), Keller (1993), Aaker (1998) e Yoo e Donthu (2001), entretanto, com uma visão mais ligada a área de marketing, o valor financeiro de uma marca é apenas o resultado da resposta do consumidor ao nome da marca (CHRISTODOULIDES e DE CHERNATONY, 2010).

Uma das primeiras definições sobre o tema abordando a visão do consumidor foi proposta no final da década de 80 por Leuthesser (1988) em seu artigo “*Defining, Measuring, and Managing Brand Equity: A Conference Summary By*” publicado pelo MSI (*Marketing Science Institute*):

Conjunto de associações e comportamentos por parte de clientes, distribuidores e empresa controladora da marca que permite a última obter maior volume ou maiores margens do que se conseguiria sem o nome da marca e que lhe dá vantagem forte, sustentável e diferenciada sobre os concorrentes. (MARKETING SCIENCE INSTITUTE, 1980 *apud* OLIVEIRA, 2013, p. 26)

Sob esse mesmo prisma, as definições propostas por Keller (1993) e Aaker (1998) destacam o consumidor como importante agente no processo de construção do valor de uma marca.

Segundo Keller (1993, p. 8) *brand equity* consiste no “efeito diferencial do conhecimento da marca na resposta do consumidor ao marketing de uma marca”. Dessa forma, pode-se dizer que uma marca tem um valor positivo (ou negativo) se o consumidor reage mais (ou menos) favoravelmente ao marketing relacionado ao

produto de uma marca conhecida do que reagiria ao mesmo produto de uma marca fictícia ou sem nome (KELLER, 1993).

Keller e Machado (2006, p.30) complementam que o “*brand equity* está relacionado ao fato de se obterem com uma marca resultados diferentes daqueles que se obteriam se o mesmo produto ou serviço não fosse identificado por aquela marca”. Logo, os efeitos de marketing exercem um forte apelo psicológico, que por sua vez influencia o consumidor na sua compra, o que por fim gera um “valor” para a marca. Keller (1998, p. 68) destaca ainda que os aspectos mercadológicos são os responsáveis por construir a equidade de uma marca. Para o autor, “*brand equity* é definida em termos dos efeitos de marketing atribuídos unicamente a uma marca”.

Aaker (1998, p. 16) também apresenta sua proposta de definição de *brand equity* e destaca sua forte relação com o nome e símbolo da marca:

O *brand equity* é o conjunto de ativos e passivos ligados a uma marca, seu nome e seu símbolo, que se somam ou se subtraem do valor proporcionado por um produto ou um serviço para uma empresa e/ou consumidores dela. Para que certos ativos e passivos determinem o *brand equity*, eles devem estar ligados ao nome e/ou símbolo da marca. Se o nome da marca ou seu símbolo for mudado, alguns ou todos os ativos ou passivos poderão ser afetados, e mesmo perdidos, embora parte deles possam ser desviados para um novo nome e símbolo. (AAKER, 1998, p. 16)

Aaker (2012) explica que esse conjunto de ativos e passivos ligados à marca pode ser dividido em quatro principais categorias: conhecimento do nome da marca, lealdade à marca, qualidade percebida e associações à marca. Dado que o *brand equity* é formado a partir desse conjunto de ativos e passivos, então o gerenciamento da equidade da marca envolve esforços para criar e melhorar esses elementos, possibilitando assim a construção de marcas fortes.

Kotler e Keller (2006) destacam como esse valor abstrato agregado à marca – que está presente na psique do consumidor – é repassado aos produtos ou serviços relacionados àquela marca:

*Brand equity* é o valor agregado atribuído a produtos e serviços. Esse valor pode se refletir no modo como os consumidores pensam, sentem e agem em relação à marca, bem como nos preços, na participação de mercado e na lucratividade que a marca proporciona à empresa. O *brand equity* é um importante ativo intangível que representa valor psicológico para a empresa. (KOTLER E KELLER, 2006, p. 270)

Por fim, o conceito apresentado pela AMA (2015) esclarece como o consumidor é influenciado por uma marca com maior *brand equity*:

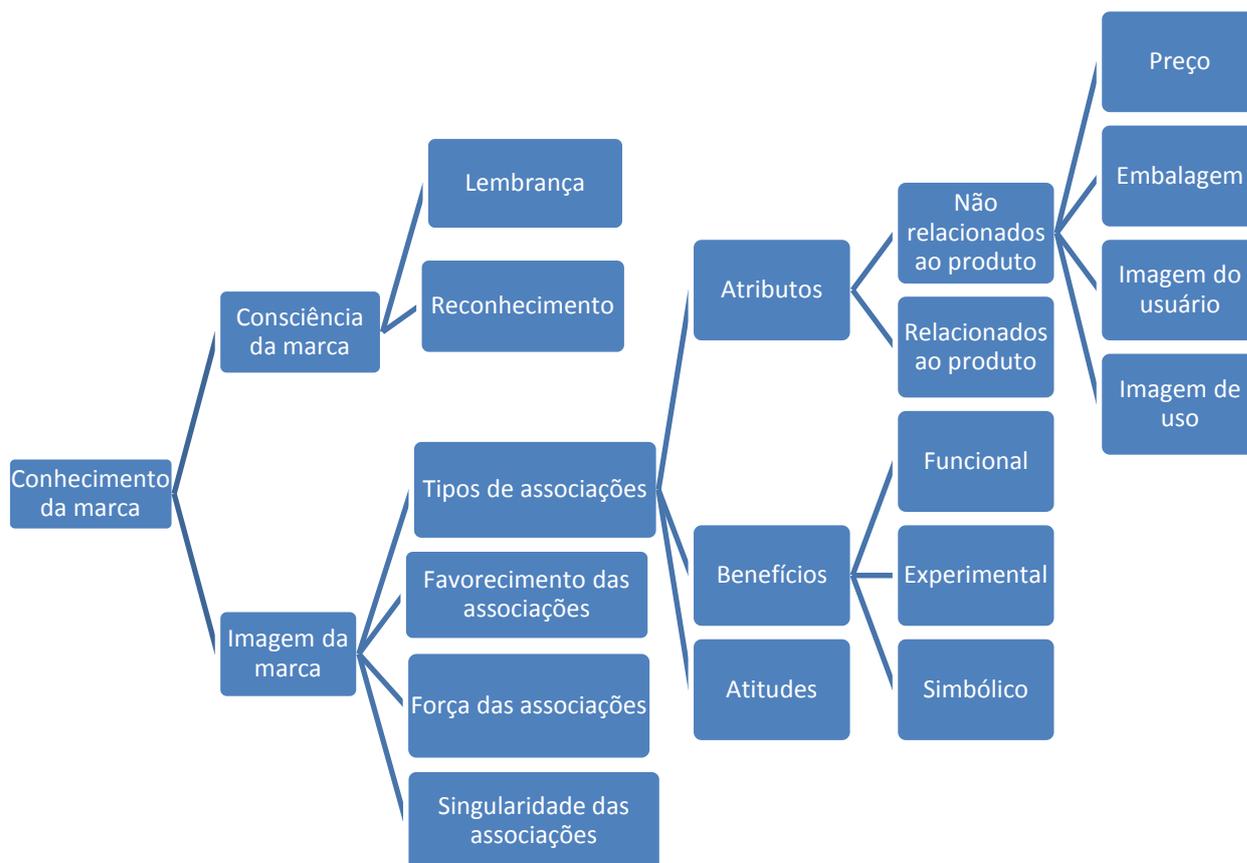
*Brand equity* é o valor de uma marca. Do ponto de vista do consumidor, *brand equity* é baseado nas atitudes do consumidor sobre atributos positivos da marca e consequências favoráveis do uso dessa marca. [...] Descreve o valor de possuir uma marca conhecida pelos consumidores, com base na ideia de que o proprietário de uma marca com prestígio entre os consumidores pode gerar mais dinheiro com a venda dos produtos relacionados a ela se comparada a produtos de uma marca com nome menos conhecido [...]. Consumidores acreditam que um produto de uma marca famosa é melhor do que produtos de marcas menos conhecida. (AMA, 2015, tradução nossa)

Apesar das definições apresentadas, evidenciando as duas principais abordagens relacionadas ao *brand equity* – com foco financeiro e baseado no consumidor – mensurá-lo não é uma tarefa trivial mesmo para os pesquisadores e profissionais da área de marketing. Para tentar alcançar esse objetivo, alguns modelos procuraram estabelecer quais são os aspectos ou dimensões que constituem a equidade de uma marca. Com base nos objetivos dessa pesquisa, serão estudados os modelos de *brand equity* com foco no consumidor, destacando-se aqueles apresentados por Keller (1993) e Aaker (1998), que são mais bem formulados na seção 2.1.3.

### 2.1.3 Valor das marcas na perspectiva do consumidor: principais modelos

Conforme a definição proposta por Keller (1993) apresentada anteriormente, a chave para construção do *brand equity* é o conhecimento da marca. O conhecimento da marca, por sua vez, pode ser conceituado como sendo “nós” na memória do consumidor onde existe um conjunto de associações relacionadas a uma determinada marca. Assim, Keller (1993) decompõe o conhecimento da marca em dois constructos distintos: consciência da marca e imagem da marca (associações). A partir dessas duas dimensões, o autor propõe seu modelo de CBBE, esquematizado na Figura 1.

FIGURA 1– DIMENSÕES DA EQUIDADE DA MARCA SEGUNDO MODELO CONCEITUAL DE KELLER



FONTE: ADAPTADO DE KELLER (1993, p. 7)

A primeira dimensão do conhecimento da marca proposta por Keller (1993) é consciência da marca; essa dimensão está relacionada à força de uma marca na memória do consumidor ou, mais especificamente, com a probabilidade de uma marca vir à mente e a facilidade com que isso ocorre. A consciência da marca pode ser construída por meio do reconhecimento da marca e recordação da marca. O reconhecimento da marca está relacionado com a habilidade de os consumidores distinguirem corretamente uma marca que eles tenham visto ou ouvido anteriormente, ao passo que a recordação da marca está relacionada a habilidade de os consumidores resgatarem em suas memórias uma marca quando lhes é fornecida uma categoria de produto (KELLER, 1993).

Keller (1993) define imagem da marca como as percepções sobre a marca mantidas através de associações na memória do consumidor; essas associações são formadas através de nós de informação ligados aos nós da marca e estabelecem o significado da marca para os consumidores. Segundo o autor, as associações à marca

podem ser analisadas sob quatro dimensões: tipos de associação, favorecimento da associação, força da associação e exclusividade da associação à marca.

Os diferentes tipos de associações à marca podem ser distinguidos através de seus níveis de abstração, ou seja, quanta informação está resumida naquela associação. São eles: atributos, benefícios e atitudes da marca (KELLER, 1993).

Os atributos são predicados que caracterizam um produto ou serviço, como seu sabor (no caso de um produto alimentício) ou nível de conforto (no caso de um carro). Os atributos podem não ser relacionados diretamente ao produto em si, mas descrevem outras características como o preço (barato/caro), à qualidade da embalagem (nesse caso não do produto propriamente dito), à imagem do usuário (que tipo de pessoa usa aquele produto) ou ainda a imagem de uso (em que tipos de situação as pessoas usam aquele produto) (KELLER, 1993).

Benefícios são os valores pessoais dos consumidores associados aos atributos de um produto ou serviço, ou seja, o que os consumidores imaginam que aquele produto ou serviço pode fazer por eles. Os benefícios podem ser funcionais, empíricos ou simbólicos. Os benefícios funcionais – que geralmente correspondem aos atributos relacionados ao produto ou serviço – são frequentemente relacionados a motivações básicas, como necessidades fisiológicas e de segurança. Benefícios empíricos relacionam-se àquilo que o consumidor sente ao utilizar o produto ou serviço – também geralmente relacionados com os atributos do produto ou serviço; satisfazem necessidades empíricas como prazeres sensoriais, diversidade e simulação cognitiva. Por fim, os benefícios simbólicos correspondem a atributos não relacionados diretamente ao produto ou serviço e atendem necessidades de aprovação social ou expressão pessoal ou ainda satisfação da autoestima (KELLER, 1993).

Atitude da marca foi incluída nos modelos de preferência do consumidor para capturar outros aspectos do produto ou serviço que não podem ser descritos em termos de atributos e benefícios, mas que influenciam na decisão do consumidor. Se os consumidores não têm motivação ou habilidade para avaliar um produto ou serviço, eles podem utilizar outros sinais externos para inferir sobre sua qualidade (KELLER, 1993).

As associações à marca construídas na mente dos consumidores sejam elas de qual tipo for – atributos, benefícios ou atitudes – podem ser classificadas de acordo com seu favorecimento, ou seja, o quão favorável à marca são essas associações. Para Keller (1993, p. 5), “o sucesso de programas de marketing é refletido na criação

de associações favoráveis à marca”. O autor explica que “o consumidor acredita tanto que a marca tem atributos e benefícios que satisfazem suas necessidades de tal modo que uma atitude positiva é formada sobre a marca” (KELLER, 1993, p. 5).

De um modo geral, os consumidores não perceberão um atributo como bom ou ruim se eles não o considerarem relevante para o produto ou serviço sendo, portanto, pouco provável que o consumidor crie uma associação favorável para um atributo que ele não considere importante (KELLER, 1993).

Vale salientar, ainda, que nem todas as associações com uma marca serão decisivas em um momento de escolha do consumidor que pode, por exemplo, associar uma marca a uma cor de embalagem específica: apesar de essa associação ser útil para localização do produto em uma prateleira, por exemplo, não é essa associação, necessariamente, que levará o consumidor a decidir qual marca escolher. Além disso, associações podem ser valiosas em algumas situações, mas menos importantes em outras. Por exemplo, atributos como velocidade e eficiência (de uma empresa de entregas, por exemplo) podem ser relevantes quando o consumidor está ansioso por sua encomenda, mas pode ter pouco impacto quando o consumidor não tem muita pressa de receber seu produto (KELLER, 1993).

Podem-se caracterizar as associações ainda de acordo com sua “força” de conexão com o “nó” da marca na memória do consumidor. Essa característica depende tanto da codificação como do armazenamento. A codificação diz respeito à forma que uma informação adentra a mente do comprador, ao passo que o armazenamento está relacionado com a forma como a associação é mantida como parte da imagem da marca. Quando um consumidor, a partir de uma ação de marketing, de fato pensa a respeito e desenvolve um significado para uma informação que se refere a um produto ou serviço de uma marca específica, mais forte são as associações criadas na memória que, por sua vez, aumentam a probabilidade de a informação ser facilmente recuperada pelo consumidor quando necessário (KELLER, 1993).

A exclusividade de associações à marca diz respeito às associações que são feitas com uma marca, mas que, obrigatoriamente, não são usufruídas pelas marcas concorrentes. Essa característica intrínseca à determinada marca é que influencia o consumidor a escolher uma marca em detrimento de outra. A construção de associações exclusivas pode ser alcançada por meio de campanhas explícitas que mostram comparações diretas com concorrentes ou de forma mais implícita, sem

atacar a marca concorrente, mas destacando qualidades que o produto da marca possui, mas que estão ausentes no produto concorrente (KELLER, 1993).

Contemporâneo de Keller (1993), que propôs seu modelo bidimensional com o objetivo de compreender a percepção do consumidor em relação à marca – capturando, assim, o *brand equity* sob a perspectiva do consumidor – Aaker (1998) sugeriu um modelo com quatro dimensões que retratam a perspectiva do consumidor, ao adicionar lealdade e qualidade percebida às duas dimensões propostas por Keller (1993). Aaker (1998) ainda expôs uma quinta dimensão, denominada “outros ativos da empresa”, relacionada estritamente à empresa, sem relação com o consumidor. As cinco dimensões do modelo de Aaker (1998) são esquematizadas na Figura 2.

**FIGURA 2 – DIMENSÕES DA EQUIDADE DA MARCA SEGUNDO MODELO CONCEITUAL DE AAKER E VALORES PROPORCIONADOS PELA EQUIDADE DA MARCA AO CONSUMIDOR E À EMPRESA**



FONTE: ADAPTADO DE AAKER (1998, p. 18)

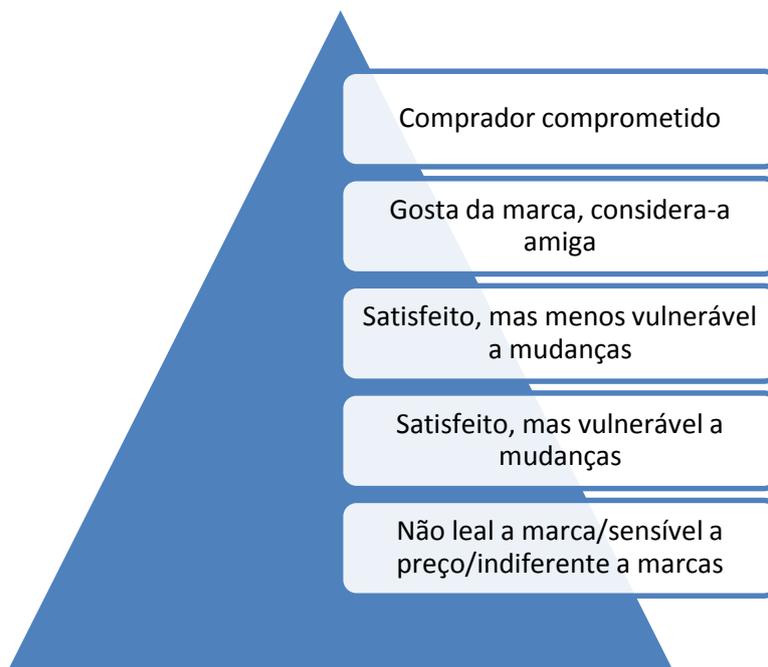
Para Aaker (1998), a lealdade do consumidor à marca é a essência do *brand equity*:

Se os consumidores lhe são indiferentes e compram segundo as características, preço e conveniência, pouco considerando o nome da marca, provavelmente há pouco *brand equity*. Se, por outro lado, continuam a comprar a marca mesmo existindo concorrentes com características superiores, preço e conveniência, existe valor na própria marca e talvez no seu símbolo e *slogans*. (AAKER, 1998, p. 40)

A lealdade a uma marca pode ser medida por meio da ligação do consumidor com a marca, de modo que quanto maior for a lealdade do consumidor àquela marca, menor a chance de trocá-la, nem mesmo ocasionalmente, por uma marca da concorrência (AAKER, 1998).

Aaker (1998) sugere uma pirâmide de lealdade para distinguir os vários níveis de lealdade a uma marca, conforme mostra a Figura 3.

**FIGURA 3 – NÍVEIS DE LEALDADE À MARCA SOB A PERSPECTIVA DO CONSUMIDOR**



**FONTE: ADAPTADO DE AAKER (1998, p. 41)**

O nível mais baixo de lealdade, representado na base da pirâmide, representa o consumidor que é completamente indiferente à marca. Aaker (1998) chama esse consumidor de “comprador mutável” ou “comprador por preço”. Para esses consumidores, o nome da marca pouco importa na hora da compra e sua escolha é feita por conveniência.

O segundo nível é formado por compradores que estão satisfeitos com o produto ou que, na pior das hipóteses, não estão insatisfeitos. São, segundo Aaker (1998), os “compradores habituais”. Embora esses consumidores não estejam a procura de alternativas, eles podem ser suscetíveis às campanhas de marcas concorrentes que apresentem um benefício novo que não seja encontrado em sua marca predileta.

O terceiro nível, composto por consumidores também satisfeitos, mas que são menos vulneráveis à mudança que o nível anterior porque temem os custos relativos a mudança, ao tempo, dinheiro ou performances. São os chamados “compradores leais”, que só trocam de marca caso sejam convencidos que os benefícios oferecidos pelo concorrente superem seus “medos” e os custos da troca (AAKER, 1998).

Os consumidores encontrados no quarto nível da pirâmide de lealdade são aqueles que realmente gostam da marca, seja com base em uma associação com o símbolo, suas experiências de utilização ou com uma alta qualidade percebida. Algumas vezes, esse sentimento do consumidor em relação à marca não pode ser nem mesmo explicado porque a relação se faz tão forte que se torna emocional. Por isso, esse grupo de consumidores é denominado “amigo” da marca (AAKER, 1998).

Por fim, no nível mais elevado da pirâmide de lealdade à marca estão os consumidores que tem “orgulho de serem descobridores ou usuários da marca. [...] A marca lhes é muito importante funcionalmente, ou como expressão do que eles são. A sua confiança é tal que eles recomendarão a marca para outras pessoas” (AAKER, 1998, p. 42). Por isso, o autor dá a eles a alcunha de “consumidores comprometidos”. Quando uma marca possui um número considerável de clientes envolvidos e comprometidos, pode-se dizer que essa é uma marca carismática.

Aaker (1998) destaca que a lealdade pode ser medida de forma direta por meio dos padrões reais de compra, sugerindo três medições possíveis: taxas de recompra, percentual de compras e número de marcas compradas. As taxas de recompra podem ser medidas pelo percentual de consumidores que compraram um produto de determinada marca e que, em sua próxima compra, voltaram a comprar um produto da mesma marca. O percentual de compras, por sua vez, pode ser medido através da seguinte pergunta: “das cinco últimas compras feitas pelo consumidor, que percentual referiu-se a cada marca comprada?”. Por fim, a lealdade a uma marca pode ser comensurada pelo número de marcas de determinado produto escolhidas em suas

últimas compras: quanto menor o número de marcas compradas, maior a lealdade do consumidor às marcas preferidas.

O autor apresenta uma ressalva: as medidas dos padrões de comportamento do consumidor podem ter limitações, ao passo que a obtenção dos mesmos pode ser inconveniente e dispendiosa, já que os dados são obtidos geralmente através de pesquisas com o consumidor.

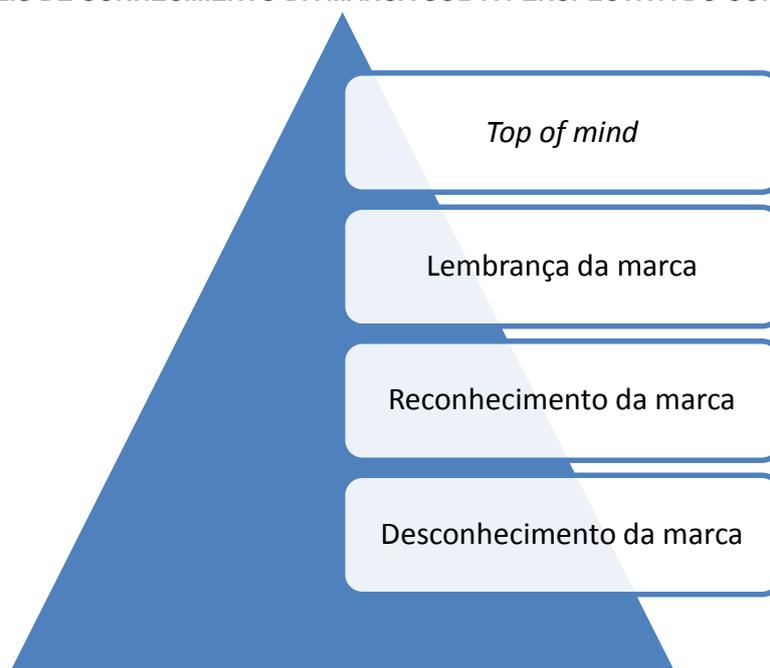
Aaker (1998, p. 45) alerta ainda que “o uso de informações de comportamento pode dificultar a discriminação daqueles que realmente mudaram de marca e as compras de múltiplas marcas por diferentes membros da família”.

Intimamente relacionado com a consciência da marca, suscitada por Keller (1993), está a segunda dimensão proposta por Aaker (1998), o conhecimento da marca. Conforme o próprio Aaker (1998, p. 64) define, “o conhecimento da marca é a capacidade que um comprador potencial tem de reconhecer ou se recordar de uma marca como integrante de uma categoria de produtos”. O autor destaca a importância dessa característica para construção do *brand equity* para o consumidor e ainda deixa clara a forte correlação entre as dimensões conhecimento da marca e associações à marca, explicando o porquê desta não existir sem aquela:

O reconhecimento da marca é o primeiro passo básico na tarefa de comunicação. Usualmente, é um desperdício tentar comunicar os atributos da marca sem que o nome esteja estabelecido o suficiente para permitir uma associação com eles. Um nome é como uma pasta de arquivo especial na mente, que pode ser preenchido com fatos, nomes e sentimentos. Sem esse arquivo disponível na memória, os fatos e sentimentos se perdem e não podem ser rapidamente acessados quando necessários. (AAKER, 1998, p. 66)

Aaker (1998) sugere que o conhecimento da marca pode ser representado por meio de três níveis bastante diferentes, conforme ilustra a Figura 4.

FIGURA 4 – NÍVEIS DE CONHECIMENTO DA MARCA SOB A PERSPECTIVA DO CONSUMIDOR



FONTE: ADAPTADO DE AAKER (1998, p. 65)

O nível mais baixo representa o total desconhecimento da marca. Nesse estágio, o consumidor sequer reconhece o nome da marca após ser estimulado com logos e símbolos ou mesmo com o próprio nome da marca (D'EMIDIO, 2009).

O próximo nível, reconhecimento da marca, é alcançado quando em um teste estimulado, os entrevistados são submetidos a um conjunto de nomes de marcas de uma determinada classe de produtos e são estimulados a identificar os nomes que eles já tenham ouvido falar antes (AAKER, 1998). O autor explica que, “embora tenha que existir um elo entre a marca e a classe do produto, o nome não precisa ser forte. Esse estágio de reconhecimento é particularmente importante para um comprador que escolhe uma marca no ponto de venda” (AAKER, 1998, p. 65), uma vez que é pouco provável que um consumidor escolha, mesmo que ocasionalmente, um produto de uma marca que ele nunca ouvira falar.

O terceiro nível na pirâmide de conhecimento de Aaker (1998) é a lembrança da marca. Diferentemente do nível anterior, a lembrança é espontânea dado que não há necessidade de estímulo. Por esse motivo, está relacionada a uma posição mais forte da marca.

No nível mais alto da pirâmide estão as marcas “*top of mind*”. Essas são as primeiras lembradas pelo consumidor quando são estimulados a pensar em marcas de uma categoria específica de produtos. Uma marca que alcança essa classificação

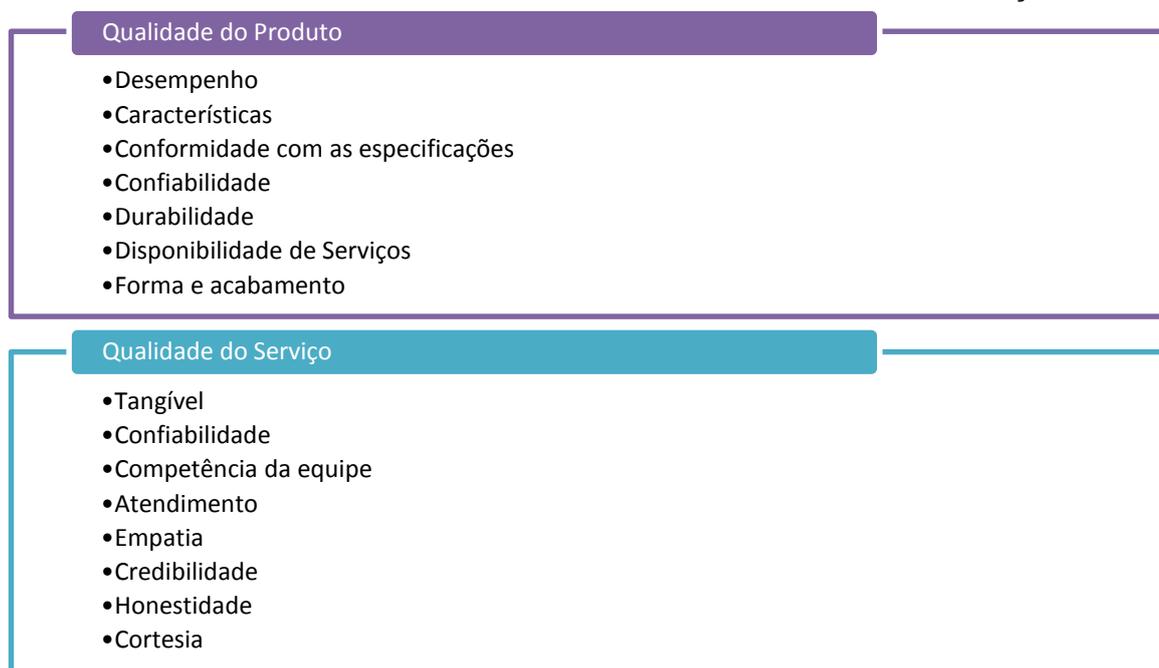
está, de forma concreta, a frente das outras marcas na lembrança do consumidor, embora possa haver outra marca bastante próxima (AAKER, 1998).

A terceira dimensão que auxilia na construção do *brand equity* é a qualidade percebida. Segundo Aaker (1998, p. 88), “a qualidade percebida pode ser definida como o conhecimento que o consumidor tem da qualidade geral ou superioridade de um produto ou serviço pretendido, em relação a alternativas”.

Cabe ressaltar que a qualidade percebida não está necessariamente ligada à qualidade real do produto – que é determinada pela qualidade dos ingredientes, do processo de fabricação, de sua fórmula ou especificações – mas sim à qualidade compreendida pelos próprios consumidores e, portanto, afetada pelos conhecimentos e julgamentos daquilo que cada um deles considera importante (AAKER, 1998).

Aaker (1998) propõe sete dimensões de qualidade associadas a produto e outras oito associadas a serviço, conforme pode ser observado na Figura 5.

**FIGURA 5 – AS QUINZE DIMENSÕES DA QUALIDADE ASSOCIADAS AO PRODUTO E AO SERVIÇO**



**FONTE: ADAPTADO DE AAKER (1998)**

O desempenho, primeira dimensão da qualidade do produto segundo Aaker (1998), tem relação com as características operacionais primárias do produto. O autor cita como exemplo o automóvel, cujas características primárias podem compreender a aceleração, direção, velocidade e conforto. Assim, um produto com bom desempenho, deve cumprir com as tarefas ou funções para as quais ele foi designado.

A qualidade percebida pode também estar associada às características de um produto que, diferentemente do desempenho, referem-se a atributos secundários do produto como, por exemplo, a luz de leitura em um automóvel. Embora tais qualidades pareçam pouco importantes a primeira vista, são elas que podem levar o consumidor a decidir pelo produto de uma marca ou de outra (AAKER, 1998).

A terceira dimensão associada à qualidade do produto percebida pelo consumidor é a conformidade com as especificações, que nada mais é que a ausência de defeitos. Aaker (1998) exemplifica que a redução da quantidade de defeitos foi fundamental para o sucesso dos fabricantes japoneses de veículo.

A confiabilidade se traduz na qualidade do desempenho do produto até o fim da vida útil daquele produto. Associada à confiabilidade e ao desempenho está a durabilidade, que reflete a vida econômica do produto (AAKER, 1998). Assim, é desejável pelo consumidor adquirir um produto que perdure por muitos anos (boa durabilidade) e que, durante esse período, possa usufruir com confiabilidade do desempenho proporcionado por tal produto.

A sexta dimensão da qualidade do produto é a disponibilidade de serviços que “reflete a capacidade de prestar serviços ao produto” (AAKER, 1998, p. 95). Desta forma, é desejável que estejam associados ao produto um serviço eficiente, competente e conveniente, o que pode gerar uma vantagem competitiva e influenciar a qualidade percebida pelo consumidor.

A sétima e última dimensão do produto, a forma e acabamento, está altamente associada à aparência ou à sensação de qualidade. Essa é uma dimensão importante porque representam uma característica que o consumidor pode facilmente julgar (AAKER, 1998).

As dimensões do serviço, por sua vez, estão muitas vezes associadas às dimensões do produto apresentadas. Grosso modo, a dimensão de competência da equipe pode, por exemplo, corresponder à dimensão de desempenho do produto, na medida em que está relacionada à entrega da função básica que está sendo adquirida pelo cliente. Nesse mesmo sentido, a dimensão tangível é similar a forma e acabamento, uma vez que ambas funcionam como sinais de competência (AAKER, 1998).

A confiabilidade, apesar do mesmo nome da dimensão do produto, refere-se a uma característica diferente. No contexto de serviços, a confiabilidade é alcançada

por meio da padronização de uma operação de serviço, possibilitando que o mesmo seja sempre executado com segurança e de forma correta (AAKER, 1998).

As demais dimensões relacionadas ao produto – atendimento, empatia, credibilidade, honestidade e cortesia – estão relacionadas, de maneira geral, com a natureza da interação entre o consumidor e a equipe de serviço (AAKER, 1998).

As associações da marca são a quarta dimensão do modelo de *brand equity* proposto por Aaker (1998) e é análoga à dimensão imagem da marca, proposta por Keller (1993) em seu *framework*. Após ter seu reconhecimento constituído, o próximo passo para promover uma marca com valor para o consumidor é conectar a ela novas associações. “Uma associação de marca é algo ‘ligado’ a uma imagem na memória” (AAKER, 1998, p. 114). Em outras palavras, associações são tudo aquilo que o consumidor relaciona com uma marca.

Esses relacionamentos, segundo o mesmo autor, possuem um grau de força, que é diretamente proporcional à quantidade de experiências ou exposições às comunicações e ações de marketing da empresa. Além disso, quando apoiadas em outras redes de associações anteriormente construídas pela marca, esses relacionamentos se tornam ainda mais fortes na memória do consumidor. Quando há, na mente do consumidor, um conjunto de associações organizadas de forma significativa, está formada, para esse cliente, a imagem da marca.

Aaker (1998) apresenta onze tipos de associação que podem ser construídas na mente do consumidor para formar a imagem de um produto. Essas associações estão representadas na Figura 6.

**FIGURA 6 – ASSOCIAÇÕES AO NOME OU SÍMBOLO DA MARCA CONSTRUÍDAS PELO CONSUMIDOR PARA FORMAÇÃO DA IMAGEM DE UM PRODUTO OU SERVIÇO**



**FONTE: ADAPTADO DE AAKER (1998)**

A estratégia de posicionamento de marca mais comum e ao mesmo tempo eficaz consiste em desenvolver na mente do consumidor uma associação do produto com um ou alguns de seus atributos. Quando esse atributo é expressivo para o consumidor, a associação criada pode ser suficiente para que ele escolha a marca relacionada (AAKER, 1998).

Aaker (1998) explica que associações com atributos intangíveis – como qualidade percebida, liderança tecnológica e outros relacionados às especificações técnicas – são análogas às anteriores, mas relacionadas a características menos “concretas” dos produtos e, portanto, mais difíceis de serem avaliadas pelo consumidor.

Os benefícios aos consumidores e os atributos do produto as vezes se confundem pelo fato de que, geralmente, um atributo proporciona um benefício ao consumidor. Aaker (1998) exemplifica que o controle de cárie é, ao mesmo tempo, uma característica do produto “creme dental”, mas também um benefício proporcionado a quem o utiliza. De modo semelhante, o banco de couro – que pode ser um atributo de um automóvel – pode proporcionar ao motorista o benefício do conforto. Segundo o autor, os benefícios podem ainda ser racionais – estritamente relacionado a um atributo do produto, fazendo parte de um processo lógico de decisão – ou psicológicos – quando se relacionam aos sentimentos motivados no ato da compra ou uso da marca.

Apesar de o preço relativo do produto ser um atributo desse produto, Aaker (1998), em concordância com Keller (1993), considera essa característica separadamente e destaca que, dado a grande quantidade de níveis de preço em determinadas classes de produto, a avaliação de uma marca nessas classes tem início na determinação de seu posicionamento em um ou dois desses níveis.

Outro tipo de associação que pode ser feita a uma marca é referente ao seu uso ou aplicação, ou seja, que tipos de ocasiões estão associados ao consumo de determinado produto (AAKER, 1998). Uma marca de café pode, por exemplo, associar seu produto ao café da manhã para dar energia para começar o dia, enquanto outra pode associar o uso de seu produto após as refeições. Naturalmente, uma marca pode associar seu produto a várias ocasiões, não sendo restrito a uma ou duas delas. Essas associações, segundo Aaker (1998), também fazem parte das estratégias de posicionamento da marca no mercado.

As associações entre as marcas e tipos de usuário são similares àquelas com o uso ou aplicação e, assim como a anterior, ajudam a determinar o posicionamento da marca, conforme explica Aaker (1998, p. 129): “outro enfoque de posicionamento é associar uma marca com um tipo de usuário do produto ou consumidor. [...] Identificar uma marca com o seu segmento-alvo é, com frequência, uma boa maneira de apelar para ele.”

Outra estratégia frequentemente utilizada pelos profissionais de marketing é planejar campanhas que relacionem uma marca a uma celebridade, de modo que o consumidor crie associações entre essas duas entidades e, não raro, transfira as associações com a celebridade para a marca (AAKER, 1998).

O estilo de vida ou a personalidade daqueles que fazem uso de uma marca também é uma associação frequentemente construída na memória dos consumidores. Conforme ilustra Aaker (1998, p. 132), “uma marca [...] pode ser impregnada de uma série de características muito similares de personalidade e de estilo de vida do consumidor, ou do seu proprietário”.

Para Aaker (1998, p. 132), “algumas marcas precisam tomar decisões críticas de posicionamento que envolvem associações de classe de produto”. O autor cita o exemplo da GE, Sony e HP que cobrem uma grande gama de produtos e que, em decorrência disso, não ganham associações com uma classe específica. Apesar disso, ainda que arriscado, marcas como essas podem desenvolver associações intangíveis, como inovação ou qualidade percebida, as quais podem ser associadas aos produtos sob seu guarda-chuva.

Aaker (1998) explica que algumas vezes é interessante para uma marca construir associações com uma marca concorrente. Tentar colocar para os consumidores uma marca como opção a uma marca líder de mercado pode ser uma estratégia eficaz.

Por fim, uma marca pode ser associada a um país ou área geográfica. Aaker (1998) mostra o exemplo da Alemanha que é frequentemente ligada a cervejas, a França, a artigos de moda e perfumes ou a Rússia, a marcas de vodka. Da mesma forma como as associações das marcas com celebridades, as associações com países possibilitam transferência de características dos países para as marcas, de modo que esses atributos passem a fazer parte da imagem da marca na memória do consumidor.

Além dos modelos propostos por Keller (1993) e Aaker (1998), outros modelos para conceituação da equidade de marca sob a perspectiva do consumidor foram propostos durante os anos 90 e 2000; dentre eles, destacam-se os trabalhos de Blackston (1992), Sharp (1996), Berry (2000) e Burmann, Jost-Benz e Riley (2009).

Blackston (1992) defende a ideia de que em um mercado dominado por marcas corporativas ou marcas “guarda-chuva”, as escolhas dos consumidores dependem muito mais da percepção global que esse comprador possui a respeito da companhia do que simplesmente a avaliação de um único produto ou serviço oferecido por essa corporação. Destarte, o autor propõe que a equidade da marca pode ser baseada em dois componentes: a confiança na marca e a satisfação do consumidor com relação a marca.

Sharp (1996), por sua vez, propõe que a equidade da marca é construída a partir da confiança na marca, da imagem da marca e do relacionamento com os clientes e possíveis franquias. Para o autor, a “consciência é um importante ativo resultante das atividades de comunicação do marketing da empresa e um reflexo de sua cota no mercado” (SHARP, 1996, p. 6, tradução nossa). A imagem da marca, em seu turno, reflete a reputação da companhia na visão dos consumidores, além de constituir um importante diferencial estratégico na medida em que dificilmente pode ser copiada por um concorrente; quando um competidor ainda assim faz opta por esse caminho, corre o risco de beneficiar o atual detentor da imagem, tornando-a ainda mais forte. O autor completa seu tripé conceitual relacionado à equidade de marca defendendo que o bom relacionamento com os clientes é parte fundamental desse constructo, afirmando que a garantia de uma clientela no futuro se dá a partir da manutenção dos clientes atuais.

Para Berry (2000), a equidade da marca é formada pela consciência da marca e pelo significado da marca. O autor define consciência da marca como “a habilidade do cliente de reconhecer e recordar da marcada quando uma deixa é fornecida” (BERRY, 2000, p. 129, tradução nossa). Por outro lado, o significado da marca tem também relação com as percepções dominantes do consumidor em relação a marca. Em outras palavras, o significado da marca pode ser definido como “uma fotografia da impressão do cliente em relação à marca e suas associações” (BERRY, 2000, p. 129, tradução nossa). Para o autor, tanto a consciência quanto o significado da marca são afetados pela maneira como a marca é apresentada pela companhia, pelas comunicações externas relacionadas à marca e, finalmente, pela experiência do consumidor com a companhia.

Burmann, Jost-Benz e Riley (2009) apresentam um modelo conceitual de equidade de marca sob a perspectiva do consumidor baseado em cinco dimensões: clareza do benefício da marca, qualidade percebida da marca, singularidade do benefício da marca, simpatia à marca e confiança na marca. Para os autores, uma clara imagem de marca só pode ser alcançada por meio de coerência e integração com as associações subjacentes à marca.

Entretanto, a clareza dos benefícios da marca, por si só, não são suficientes para a criação de uma marca forte, de modo que a qualidade da marca percebida pelos consumidores também precisa ser considerada:

Esse indicador amplamente aceito implica no potencial de desempenho da marca de acordo com a percepção do consumidor. Portanto, o indicador está primariamente relacionado às associações aos atributos da marca. Essa variável representa associações benéficas à marca, tanto funcionais como simbólicas, na medida em que a qualidade por si só não representa em si um fim, mas sim um meio pelo qual as necessidades dos compradores são satisfeitas. Sugere-se uma operacionalização bastante genérica desse atributo, utilizando-se uma medida de qualidade de atributo único (BURMANN, JOST-BENZ e RILEY, 2009, p. 394, tradução nossa).

A singularidade do benefício da marca, por sua vez, refere-se ao grau como marca é notada como diferente em relação a seus concorrentes; quanto maior esse grau, maior a originalidade da marca; já a simpatia à marca “é a medida do nível da percepção positiva da marca [...] refletindo uma sintonia fina entre a percepção da imagem da marca na mente do consumidor e a identidade da marca comunicada pela companhia” (BURMANN, JOST-BENZ e RILEY, 2009, p. 394, tradução nossa); essa dimensão é complementada pela confiança na marca: ao passo que a primeira é denotada pela preferência espontânea e imediata pela marca, a segunda é resultado de um longo processo de avaliação cognitiva (BURMANN, JOST-BENZ e RILEY, 2009).

Por meio dos modelos conceituais apresentados, torna-se possível compreender quais aspectos são levados em consideração para análise do *brand equity* sob o ponto de vista do consumidor. Entretanto, conforme os objetivos definidos para essa pesquisa, a análise do *brand equity* baseado no consumidor deverá ser feita no contexto das redes sociais. Sendo assim, na seção 2.2, são apresentados os principais fundamentos relacionados a esse tema.

## 2.2 REDES SOCIAIS

Os estudos de padrões de amizade em pequenos grupos realizados em 1934 por Jacob Moreno (MORENO, 1934) marcaram o início da sociometria, considerada como precursora da análise de redes sociais e da psicologia social (WASSERMAN e FAUST, 1994; OLIVEIRA, 2011), mas foi apenas em 1997 que as redes sociais ganharam o mundo virtual, com o advento do primeiro *site* de rede social, chamado SixDegrees.com (ELLISON, 2007). Nesse período, foram desenvolvidas bases teóricas para a análise das redes sociais, cujas principais definições e elementos serão investigados nas seções 2.2.1 a 2.2.5. Em seguida, na seção 2.2.6, são apresentadas as definições referentes às redes sociais virtuais.

### 2.2.1 Análise de redes sociais

A base teórica para análise de redes sociais é resultante do esforço conjunto de várias áreas do conhecimento, como ciências sociais, matemática, estatística e computação, sendo os pesquisadores da área de sociologia, psicologia social e antropologia os pioneiros nesse campo de pesquisa (WASSERMAN e FAUST, 1994; FREEMAN, 1996; BORGATTI, 2002).

Guiado pelo interesse em estudar o comportamento humano tanto social como psicologicamente, sobretudo a dinâmica de grupos, Jacob Moreno propôs em 1934 um modo de representar graficamente o relacionamento interpessoal em um grupo. Nesse diagrama, denominado sociograma, as pessoas eram simbolizadas por pontos distribuídos em um espaço bidimensional, e as relações entre essas unidades sociais eram reproduzidas por meio de linhas que conectavam pares desses pontos. O reconhecimento de que sociogramas carregavam em si um grande potencial para o estudo das estruturas das redes sociais levou rapidamente ao desenvolvimento de técnicas analíticas, ao mesmo tempo em que se percebeu que as matrizes também poderiam ser utilizadas para a representação de dados das redes sociais (WASSERMAN e FAUST, 1994).

Apesar de Jacob Moreno ter sido o primeiro a ter estudado a inter-relação entre pessoas, foi Barnes (1954) quem cunhou o termo “rede social” para representar o conceito de uma rede de relações entre entidades sociais, que podem ser – mas que não se restringem a – indivíduos, organizações ou objetos. Assim, a análise de redes sociais tem por objetivo compreender as relações entre essas entidades sociais e as implicações desse relacionamento (WASSERMAN e FAUST, 1994).

Logo, é necessário esclarecer alguns elementos fundamentais para a análise de redes sociais, os quais são apresentados na subseção 2.2.2.

### 2.2.2 Elementos fundamentais

*Atores* são entidades sociais, podendo ser um ente coletivo, corporativo ou individual (FREITAS e PEREIRA, 2004), dotado de um conjunto de características ou atributos que “tipificam seus comportamentos e objetivos, exercendo influência também na conectividade desse ator dentro da rede” (SANTOS, 2010, p. 17). Os atores podem ser todos do mesmo tipo, formando redes denominadas *único-modo*,

ou podem ser de tipos diferentes, constituindo uma rede *composta* (WASSERMAN e FAUST, 1994). Um aluno de uma turma, um curso ou um departamento dentro de uma Universidade, carros no trânsito ou países são todos exemplos de atores. A relação entre alunos pode ser estruturada por meio de uma rede único-modo, enquanto que a relação de pertencimento entre alunos e cursos constitui uma rede composta.

*Laços de relacionamento* ligam os atores uns aos outros, sendo também conhecidos como laços sociais ou, simplesmente, ligações (WASSERMAN e FAUST, 1994). Os tipos mais comuns de ligação são a avaliação pessoal (amizade, respeito), transferência de recursos materiais (compra e venda), transferências de recursos não materiais (comunicação, envio e recebimento de informação), movimento (deslocamento de pessoas, migração, troca de lugares), familiar (casamento e descendência) e, por fim, física (pontes, rios, estradas) (SANTOS, 2010). As ligações podem ser classificadas quanto ao direcionamento (direcionais ou não-direcionais) e valoração (podendo ser dicotômicas, quando não assumem valores, ou valoradas, podendo assumir valores discretos ou contínuos) (OLIVEIRA, 2011). A *relação* é um “conjunto de ligações de um tipo específico entre membros de um grupo” (WASSERMAN e FAUST, 1994, p. 20, tradução nossa).

O conjunto formado por uma ligação e dois atores é conhecido como *díade*; de modo análogo, o conjunto formado por três atores e suas possíveis ligações é chamado de *tríade*. De forma mais abrangente, um *subgrupo* consiste no conjunto de  $n$  atores e suas respectivas ligações (WASSERMAN e FAUST, 1994). Oliveira (2011, p.42) explica que “a análise de díades busca identificar [...] se os laços são recíprocos e se, em uma rede multi-relacional, um conjunto específico de relações múltiplas tende a ocorrer simultaneamente”.

Quando o conjunto de atores e suas ligações pode ser delimitado ou descrito por alguma característica comum, obtém-se o que os teóricos da área de análise de redes sociais chamam de *grupo* (OLIVEIRA, 2011). De modo mais sistemático, pode-se dizer que um grupo “consiste em um conjunto finito de atores que por razões conceituais, teóricas ou empíricas, são tratados como um conjunto finito de indivíduos sobre os quais medidas de uma rede podem ser realizadas” (WASSERMAN e FAUST, 1994, p. 19, tradução nossa).

Em posse desses conceitos, é possível agora definir de modo mais sistemático o que é uma rede social. Wasserman e Faust (1994, p. 20) definem uma rede social como “um conjunto finito ou conjuntos finitos de atores e a relação ou relações

determinadas entre eles”. É exatamente a presença da informação relacional que diferencia as redes sociais de outros tipos de dados, como um conjunto de registros tradicionais, por exemplo, nos quais existem atores (objetos), mas não há um relacionamento entre essas entidades.

Assim, a análise de redes sociais tem como objetivo detectar e interpretar padrões de relacionamentos sociais que ocorrem entre os atores dessa rede (NOOY, MRVAR e BATAGELJ, 2011). Para que medidas para análise de redes sociais possam ser realizadas, métodos matemáticos formais devem ser utilizados para representação das redes sociais, sendo que grafos (teoria dos grafos) e as matrizes (álgebra matricial) são as formas mais comuns de estruturação (CASTRO, 2005; FRANÇA, 2012; CAMPOS, 2014), sendo apresentadas nas subseções a seguir.

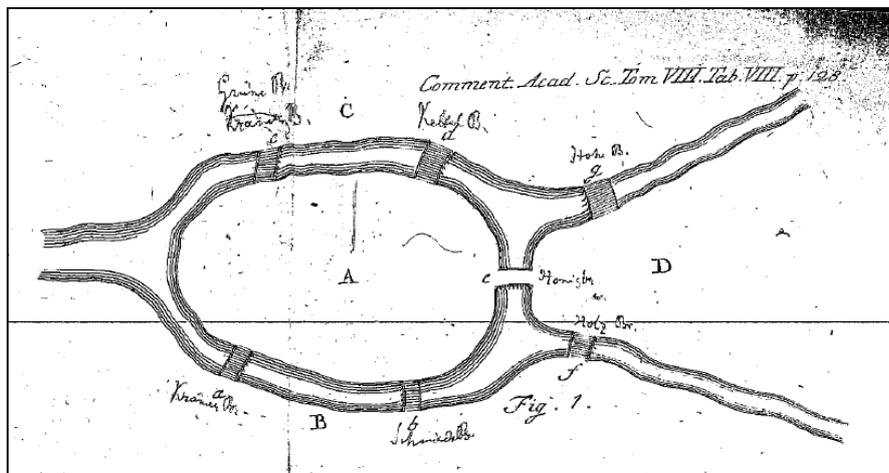
### 2.2.3 Representação por meio de grafos

Os grafos são conhecidos na sociologia como sociogramas (CASTRO, 2005), dado que a primeira proposta de uma representação gráfica de uma rede social foi proposta por Jacob Moreno, em 1934 (WASSERMAN e FAUST, 1994).

Entretanto, dois séculos antes, em 1736, Leonhard Euler escreveu seu artigo "*Solutio Problematis ad Geometriam Situs Pertin*" (ou, em uma tradução livre para o português, "A Solução de um problema relacionado à geometria de posição"), fato que levou muitos autores a atribuir à Euler o mérito pela gênese da teoria dos grafos (BARNES e HARARY, 1983; ALEXANDERSON, 2006; HSU e LIN, 2008). Em seu trabalho, Euler resolveu uma questão bastante discutida pelos habitantes da cidade de Königsberg: havia um caminho fechado – ponto de saída igual ao ponto de chegada – que cruzava exatamente uma vez cada uma das sete pontes sobre o rio Pregel?

A Figura 7, extraída do artigo original de Euler, esquematiza o problema.

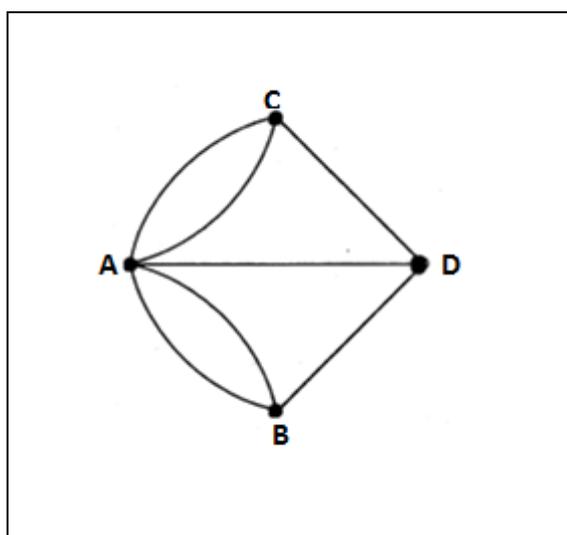
**FIGURA 7 – ESQUEMATIZAÇÃO PROPOSTA POR EULER PARA RESOLUÇÃO DO PROBLEMA DAS SETE PONTES DE KONIGSBERG**



**FONTE: EULER (1736)**

O rio Pregel possuía vários braços, dividindo a cidade em quatro áreas: a ilha de Kneiphof (A) e as regiões de Vorstadt (B), Altstadt (C) e Löbnicht (D). Para resolver o problema, Euler desenhou uma simples estrutura denominada grafo, no qual os vértices representavam as quatro áreas (A, B, C e D) que, por sua vez, eram ligadas por arestas, as quais representavam as sete pontes. A estrutura pode ser representada pela Figura 8:

**FIGURA 8 – REPRESENTAÇÃO DO PROBLEMA DAS SETE PONTES DE KONIGSBERG ATRAVÉS DE UM GRAFO**

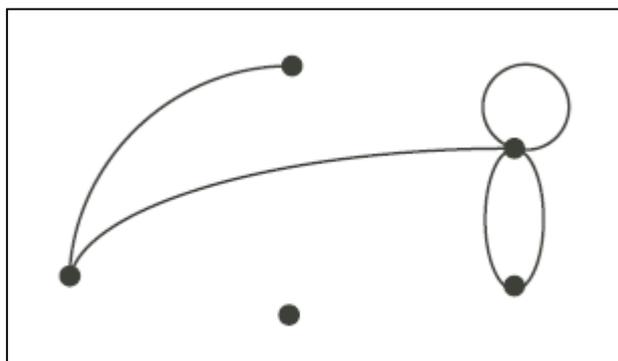


**FONTE: ADAPTADO DE BONDY E MURTY (1976, p. 51)**

Em seu artigo, Euler demonstra que é impossível cruzar as sete pontes em uma jornada única, argumentando que o grafo deve ser conectado e em cada um dos vértices deve incidir um número par de arestas (GRIBKOVSKAIA, HALSKAU e LAPORTE, 2007). A partir dessa demonstração de Euler, desenvolveram-se as bases teóricas relacionadas aos grafos que são apresentadas a seguir.

Conceitualmente, um grafo é formado por um conjunto de elementos (denominados vértices ou nós) e suas relações (denominadas arestas, as quais conectam pares de vértices) (LÉZORAY e GRADY, 2012; RUOHONEN, 2013), conforme demonstra a Figura 9.

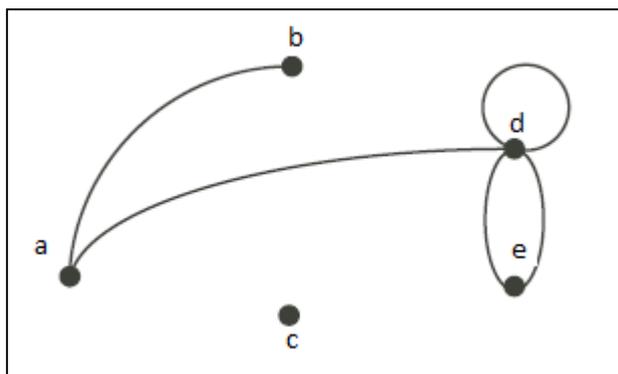
**FIGURA 9 – EXEMPLO DE UM GRAFO FORMADO POR UM CONJUNTO DE VÉRTICES E ARESTAS**



**FONTE: RUOHONEN (2013, p.1)**

Formalmente, um grafo  $G$  consiste em um conjunto  $V$ , finito e não-vazio, de  $p$  vértices, unidos por um conjunto  $E$  de  $q$  pares não-ordenados de distintos pontos de  $V$ . Cada par  $e = \{u, v\}$  de vértices em  $E$  é denominado aresta de  $G$ , e diz-se que  $x$  liga  $u$  e  $v$ . Escreve-se  $e = uv$  e diz-se que  $u$  e  $v$  são pontos adjacentes (pode ser denotado por  $u \text{ adj } v$ ); diz-se que o vértice  $u$  e a aresta  $e$  são incidentes entre si e que o ponto  $u$  é uma terminação de  $e$ ; analogamente,  $v$  e  $e$  são incidentes e  $v$  é uma terminação de  $e$ . Se duas arestas distintas  $e$  e  $f$  são incidentes com um ponto em comum, então se diz que são arestas adjacentes. Um grafo com  $p$  vértices e  $q$  arestas é chamado de grafo  $(p, q)$ . Assim, um grafo  $(1, 0)$  é chamado de grafo *trivial* (HARARY, 1969). A Figura 10 mostra a representação formal de um grafo  $(5, 5)$  formado pelo conjunto de vértices  $V = \{a, b, c, d, e\}$  e pelo conjunto de arestas  $E = \{(a, b), (a, d), (d, d), (d, e), (d, e)\}$ , podendo também ser representado por  $G(V, E)$ .

FIGURA 10 – REPRESENTAÇÃO FORMAL DE UM GRAFO

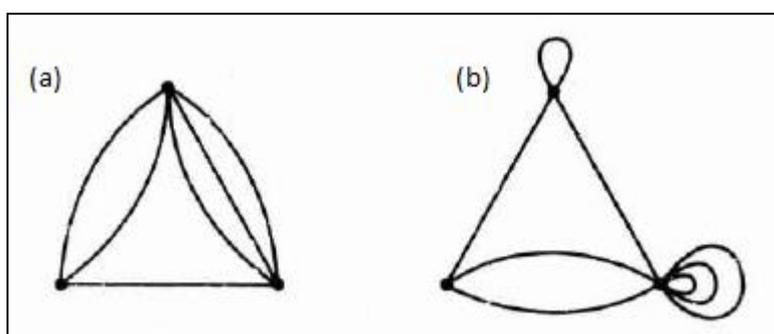


FONTE: ADAPTADO DE RUOHONEN (2013, p.1)

A maior parte das definições e conceitos na teoria dos grafos é sugerida pela própria representação gráfica dos elementos. Assim, uma aresta com terminações idênticas é chamada de *loop* (exemplo Figura 10, aresta (d,d)), ao passo que uma aresta com diferentes terminações é chamada de *ligação* (exemplo Figura 10, aresta (a,b)). Um grafo (1,0), ou seja, com um vértice e nenhuma aresta é chamado de *trivial* e os demais de *não-trivial*. Um grafo é dito *simple* quando não possui loops e não há mais de uma ligação entre dois vértices distintos (BONDY e MURTY, 1976).

Os grafos podem receber nomes especiais dadas suas características. Em um *multigrafo*, não há *loops*, mas mais de uma aresta pode unir dois diferentes vértices. Essas arestas são chamadas *paralelas*. Esse tipo de grafo também é conhecido como *múltiplas arestas*. Se tanto loops quanto múltiplas arestas são permitidos, tem-se um *pseudografo* (HARARY, 1969). A Figura 11 mostra exemplos de um multigrafo (a) e um pseudografo (b). O grafo da Figura 8, relativo ao problema de Königsberg, também é um exemplo de multigrafo.

FIGURA 11 – EXEMPLOS DE REPRESENTAÇÃO DE UM (A) MULTIGRAFO E (B) UM PSEUDOGRAFO



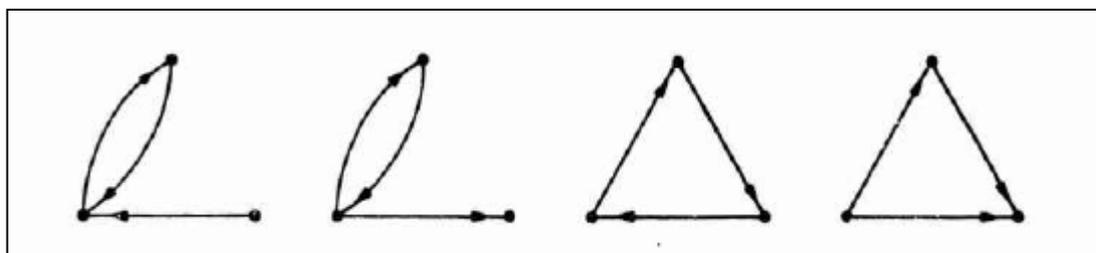
FONTE: HARARY (1969, p. 10)

Harary (1969) apresenta também os conceitos de *grafo direcionado* e *grafo orientado*:

Um *grafo direcionado* ou *dígrafo*  $D$  consiste em um conjunto finito e não-vazio  $V$  de vértices unidos por uma coleção  $X$  de pares ordenados de pontos distintos. Os elementos de  $X$  são *arestas direcionadas* ou *arcos*. Por definição, um dígrafo não possui loops nem arcos múltiplos. Um *grafo orientado* é um dígrafo que não possui pares simétricos de arestas direcionadas (HARARY, 1969, p. 10, tradução nossa).

Na Figura 12, podem-se observar quatro dígrafos, com três vértices e três arcos cada, sendo que os dois últimos também podem ser chamados por grafos orientados.

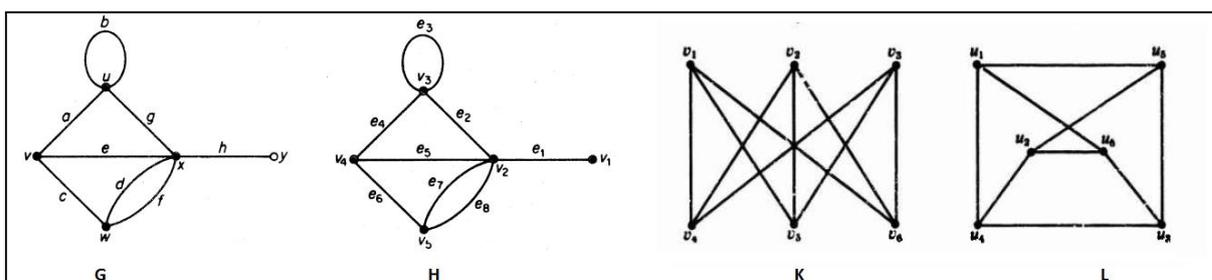
FIGURA 12 – EXEMPLOS DE DÍGRAFOS E GRAFOS ORIENTADOS



FONTE: HARARY (1969, p. 10)

Os grafos  $G$  e  $H$  são ditos *idênticos* ( $G=H$ ) se possuem o mesmo conjunto de vértices e arestas, incluindo seus rótulos. Quando os grafos são diferenciados apenas pelos rótulos, mas ainda possuem o mesmo conjunto de vértices e arestas, são chamados de *grafos isomórficos* ( $G \approx H$ ) (BONDY e MURTY, 1976). Na Figura 13, os grafos  $G$  e  $H$  são isomórficos, mas não idênticos (mesmos vértices e arestas, mas rótulos diferentes) ao passo que os grafos  $K$  e  $L$  são idênticos (mesmos vértices, arestas e rótulos).

FIGURA 13 – GRÁFOS IDÊNTICOS E ISOMÓRFICOS

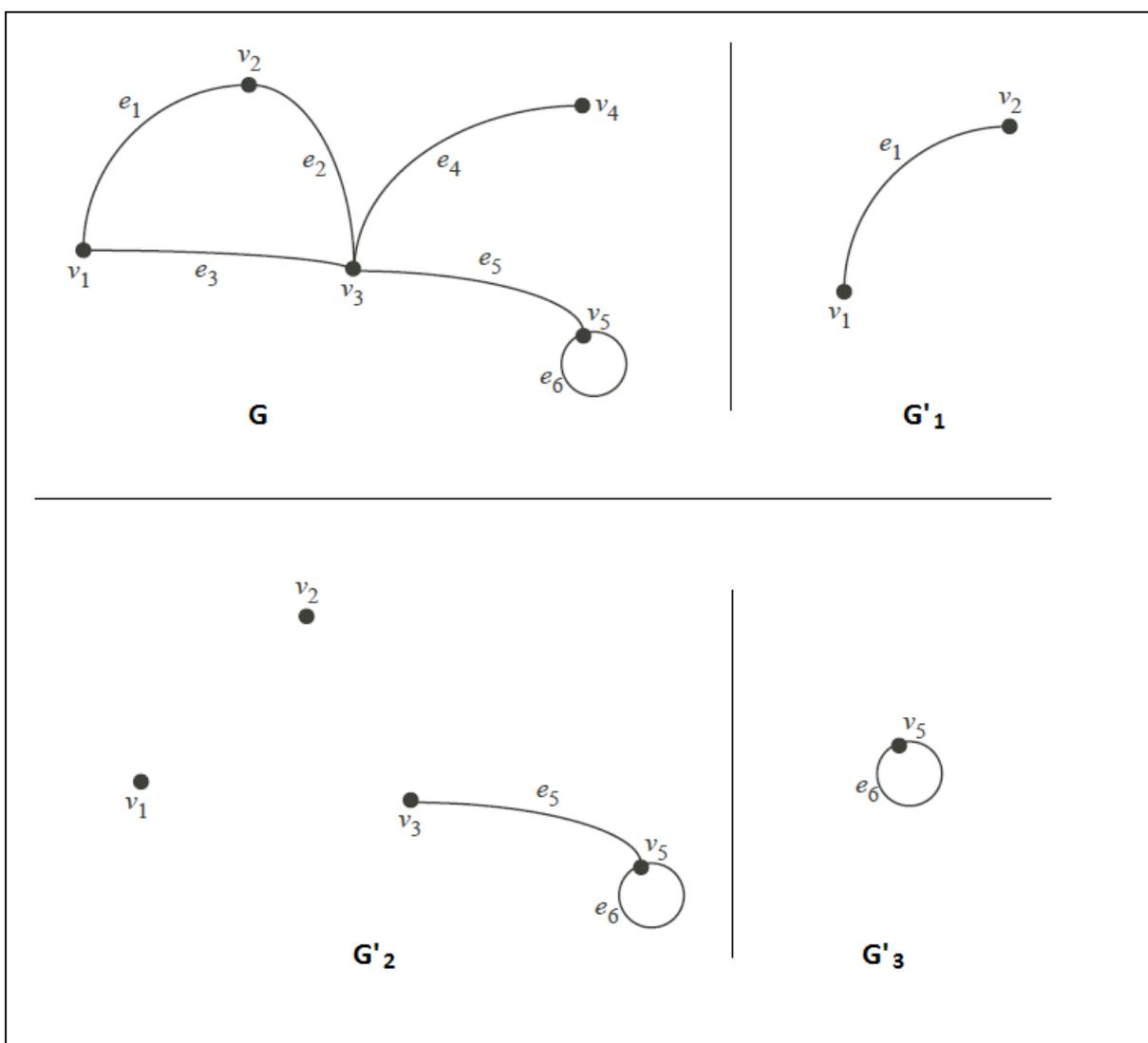


FONTE: ADAPTADO DE HARARY (1969, p. 11) e BONDY (1976, p.2)

Ressalta-se que os grafos K e L, além de idênticos, também são isomórficos, mesmo que a distribuição dos vértices e arestas no espaço seja diferente. Quando dois grafos são idênticos, eles possuem a mesma matriz de adjacência, que será apresentada na Seção 2.2.4.

Um grafo  $G'(V',E')$  é denominado *subgrafo* de  $G(V,E)$  se  $V' \subseteq V$  ( $V'$  está contido em  $V$  ou  $V'$  é um subconjunto de  $V$ ) e cada aresta  $e$  de  $G$  é também uma aresta de  $G'$  (RUAHONEN, 2013). A Figura 14 mostra um grafo  $G$  e alguns de seus possíveis subgrafos.

FIGURA 14 – APRESENTAÇÃO DO GRAFO  $G$  SEUS SUBGRAFOS  $G'1$ ,  $G'2$  E  $G'3$



FONTE: ADAPTADO DE RUOHONEN (2013, p.4-5)

Os grafos modelam relacionamentos entre nós; são esses relacionamentos que determinam se um par de nós está conectado (LÉZORAY e GRADY, 2012). O

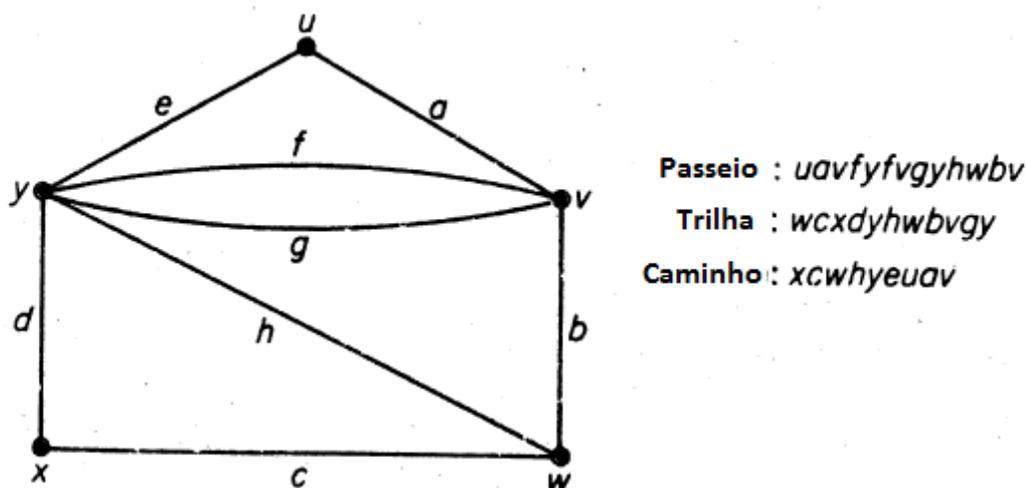
conceito de conectividade e a noção de ideias a ele relacionadas, como caminhos e árvores, são apresentados.

Um *passeio* (*walk*) por um grafo  $G$  consiste em uma sequência alternada de vértices e arestas, sempre iniciando e terminando em um vértice. Os vértices inicial e final podem ser denotados por  $v_0$  e  $v_n$ , respectivamente. Assim, um passeio liga  $v_0$  a  $v_n$  e pode ser denotado por  $v_0, e_0, v_1, e_1, v_2, e_2, \dots, v_n$ , podendo ser chamado também como um passeio  $v_0-v_n$  (HARARY, 1969). Um passeio é *aberto* se os vértices inicial e final são diferentes; caso contrário, diz-se que o passeio é *fechado* (RUAHONEN, 2013).

Um *passeio* pode ser considerado uma *trilha* (*trail*) se cada aresta é percorrida no máximo uma vez (arestas são distintas). Assim, o número máximo de vezes que um par de vértices  $u, v$  pode aparecer consecutivamente em uma trilha é igual ao número de arestas paralelas que conectam  $u$  e  $v$  (RUAHONEN, 2013).

A trilha pode ser chamado de *caminho* (*path*) se cada vértice da trilha é visitado no máximo uma vez (vértices são distintos), com exceção feita ao *caminho fechado*, no qual as terminações são coincidentes ( $v_0=v_n$ ) e portanto, visitada mais de uma vez (no início e no fim do caminho) (RUAHONEN, 2013). Um caminho fechado também pode ser chamado de *ciclo* (*cycle*) (KAVEH, 2013). A Figura 15 mostra um exemplo de passeio, trilha e caminho para o grafo apresentado  $G$ .

FIGURA 15 – OS CONCEITOS DE PASSEIO, TRILHA E CAMINHO EM UM GRAFO

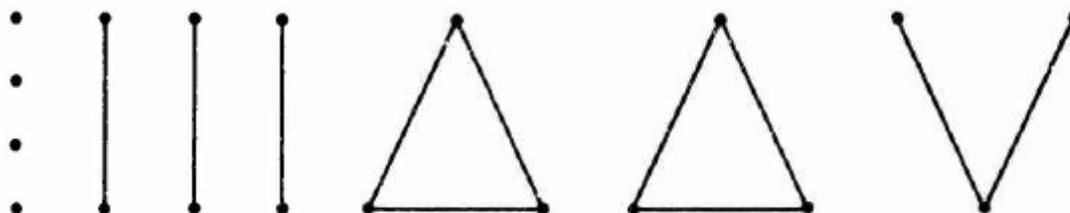


FONTE: ADAPTADO DE BONDY e MURTY (1976, p. 12)

Dois vértices são ditos *conectados* se existe pelo menos um caminho por meio do qual eles podem ser unidos. Analogamente, um grafo é *conectado* se todos os

pares de vértices são conectados. O máximo subgrafo conectado de  $G$  é chamado *componente* de  $G$ . Logo, um *grafo desconectado* possui pelo menos dois componentes (KAVEH, 2013). A Figura 16 mostra um grafo desconectado com dez componentes.

FIGURA 16 – EXEMPLO DE UM GRAFO DESCONECTADO COM DEZ COMPONENTES



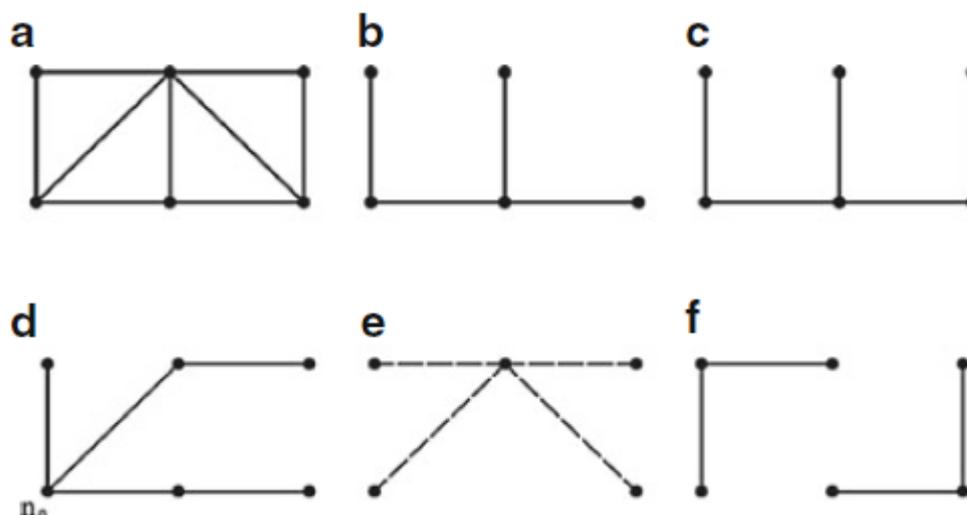
FONTE: HARARY (1969, p. 14)

O *comprimento (length)* de um passeio é dado pelo número de arestas no percurso; o *perímetro(girth)* de um grafo é o comprimento do ciclo mais curto de  $G$ , caso exista; a *circunferência (circumference)* é o comprimento do ciclo mais longo (HARARY, 1969).

A *distância  $d(u,v)$*  entre dois vértices  $u$  e  $v$  em um grafo  $G$  é o comprimento do caminho mais curto que liga  $u$  a  $v$ . Esse caminho é chamado de *geodésico* e sua distância, de *distância geodésica*. Se não existe um caminho entre  $u$  e  $v$ , então a distância entre esses vértices é infinita. O *diâmetro* de um grafo conectado  $G$  é sua maior distância geodésica (HARARY, 1969).

Uma *árvore (tree)*  $T$  de um grafo  $G$  consiste em um subgrafo conectado de  $G$  que não possui ciclo. O conjunto de árvores de  $G$  forma uma *floresta (forest)*. Se uma árvore possui todos os nós de  $G$ , diz-se que é uma *árvore abrangente (spanning tree)*. O complemento de  $T$  em  $S$  é denominado *co-árvore (cotree)*, representado por  $T^*$ . Os elementos de  $T$  são chamados *galhos (branches)* e os de  $T^*$ , *cordas (chords)*. Uma *árvore de rota curta (shortest route tree)*, enraizada em um vértice específico de  $G$ , é uma árvore na qual a distância entre cada vértice de  $T$  e seu nó-raiz é mínima (KAVEH, 2013). A Figura 17 apresenta exemplos desses conceitos.

FIGURA 17 – ALGUNS EXEMPLOS DE ÁRVORES FORMADAS A PARTIR DO GRAFO (A)



FONTE: KAVEH (2013, p. 20)

Em (a), é apresentado o grafo  $G(6,9)$ ; em (b), um exemplo de uma árvore  $T$  a partir de  $G$ ; em (c) tem-se um exemplo de uma árvore abrangente de  $G$ ; em (d) é demonstrada uma árvore de rota curta enraizada em  $n_0$ ; em (e) exemplifica-se uma co-árvore de  $G$ ; por fim, (f) mostra uma floresta com duas árvores de  $G$ .

Além dos grafos, as redes sociais podem ser também representadas por meio de matrizes. Essa forma é apresentada na seção 2.2.4.

#### 2.2.4 Representação por meio de matrizes

A complexidade e velocidade de execução de algoritmos de manipulação de grafos estão intrinsecamente relacionadas com a estrutura de dados utilizada para representação desses diagramas. As matrizes têm sido as estruturas de dados mais comumente utilizadas pelos computadores para representação de grafos, ao prover um eficiente mecanismo de armazenamento, sobretudo para grafos esparsos (CASTRO, 2005; LÉZORAY e GRADY, 2012). A utilização de matrizes como estrutura de dados para representação dos grafos possibilita ainda que muitos conceitos de álgebra matricial possam ser relacionados àqueles apresentados pela teoria dos grafos (KAVEH, 2013). Os tipos mais comuns de matrizes utilizadas para representação dos grafos são: matriz de incidência, matriz de adjacência e matriz laplaciana (HARARY, 1969; LÉZORAY e GRADY, 2012; KAVEH, 2013; RUOHONEN, 2013).

Uma *matriz de incidência* para o grafo  $G$  é uma matriz  $B=[b_{ij}]_{n \times m}$  na qual os vértices são rotulados de 1 a  $n$  e as arestas de 1 a  $m$ , sendo definida por Kaveh (2013):

$$\begin{cases} b_{ij} = 1, & \text{se o vértice } i \text{ e a aresta } j \text{ são incidentes} \\ b_{ij} = 0, & \text{caso contrário} \end{cases} \quad (1)$$

Por sua vez, a *matriz de adjacência* para o grafo  $G$  é uma matriz  $A=[a_{ij}]_{n \times n}$  na qual os vértices são rotulados de 1 a  $n$ , sendo definida por Kaveh (2013):

$$\begin{cases} a_{ij} = 1, & \text{se o vértice } i \text{ e } j \text{ são adjacentes} \\ a_{ij} = 0, & \text{caso contrário} \end{cases} \quad (2)$$

Uma *matriz de graus* é uma matriz diagonal  $D=[d_{ij}]_{n \times n}$  que contém os graus de cada vértice, dada por Kaveh (2013):

$$\begin{cases} d_{ij} = \text{grau do vértice } i, & \text{se } i = j \\ d_{ij} = 0, & \text{caso contrário} \end{cases} \quad (3)$$

Por fim, a matriz laplaciana pode ser obtida pela subtraindo-se a matriz de adjacência da matriz de graus (KAVEH, 2013):

$$L = D - A \quad (4)$$

A Figura 18 mostra as matrizes de incidência, adjacência, grau e laplaciana, obtidas a partir do grafo  $G$ .

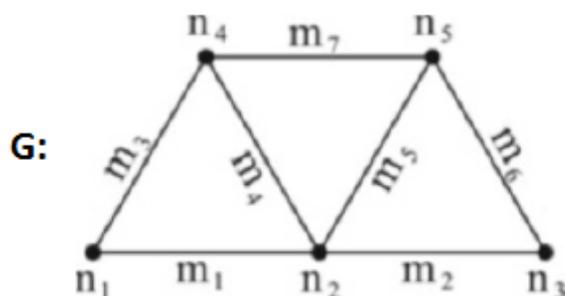
FIGURA 18 – MATRIZES COMO FORMA REPRESENTAÇÃO DOS GRAFOS

$$\mathbf{B} = \begin{matrix} & m_1 & m_2 & m_3 & m_4 & m_5 & m_6 & m_7 \\ \begin{matrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \end{matrix} & \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \end{matrix}$$

$$\mathbf{A} = \begin{matrix} & n_1 & n_2 & n_3 & n_4 & n_5 \\ \begin{matrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} \end{matrix}$$

$$\mathbf{D} = \begin{matrix} & n_1 & n_2 & n_3 & n_4 & n_5 \\ \begin{matrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \end{matrix} & \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix} \end{matrix}$$

$$\mathbf{L} = \begin{matrix} & n_1 & n_2 & n_3 & n_4 & n_5 \\ \begin{matrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \end{matrix} & \begin{bmatrix} 2 & -1 & 0 & -1 & 0 \\ -1 & 4 & -1 & -1 & -1 \\ 0 & -1 & 2 & 0 & -1 \\ -1 & -1 & 0 & 3 & -1 \\ 0 & -1 & -1 & -1 & 3 \end{bmatrix} \end{matrix}$$



FONTE: KAVEH (2013, p. 28).

A utilização das matrizes aqui apresentadas torna possível a identificação de muitas propriedades dos grafos por meio de algoritmos computacionais, desde o grau dos vértices até o número de *spanning trees* presentes em um grafo rotulado (HARARY, 1969). As matrizes também podem ser utilizadas como ponto de partida para o cálculo de métricas relacionadas a redes sociais, que serão apresentadas a seguir.

### 2.2.5 Métricas para análise de redes sociais

A análise quantitativa por meio de métricas torna possível a comparação entre redes sociais, a identificação de mudanças na rede em função do tempo e ainda a determinação da posição relativa de indivíduos ou grupos dentro de uma rede. Assim,

as métricas para análise de redes podem ser utilizadas para identificar características da rede como um todo – densidade e centralização – ou analisar um vértice específico da rede – grau de centralidade, centralidade de intermediação e centralidade de proximidade (HANSEN, SHNEIDERMAN e SMITH, 2010).

A *densidade (density)* é uma métrica de agregação utilizada em análise de rede social aplicada para descrever o nível de interconectividade dos vértices da rede, calculada por meio da relação entre número de ligações observadas na rede e o número total de relacionamentos em potencial (HANSEN, SHNEIDERMAN e SMITH, 2010). Assim, a densidade de uma rede  $R$  com  $N$  vértices,  $E$  ligações e  $N * (N - 1)$  relacionamentos possíveis é dada por (CASTRO, 2005):

$$d(R) = \frac{E}{N * (N - 1)} \quad (5)$$

A densidade é um importante ferramental para captura de ideais sociológicos, como coesão, solidariedade e filiação (HANSEN, SHNEIDERMAN e SMITH, 2010), além de ser um importante indicador do potencial da rede em difundir informação entre seus nós (HANNEMAN e RIDDLE, 2005). Diz-se que a rede é *densa* quando  $d(R)$  se aproxima de 1 ou *esparsa* quando apresenta baixa densidade –  $d(R)$  mais próximo de 0 (WASSERMAN e FAUST, 1994).

A *centralização (centralization)* é também uma métrica de agregação que reflete se uma rede é centrada em um nó – ou em poucos importantes nós – ou se é uma rede mais distribuída, ou seja, pouco concentrada. Redes com alto nível de centralização geralmente possuem muitas arestas que emanam de um pequeno conjunto de importantes vértices, ao passo que redes descentralizadas apresentam uma baixa variação em relação ao número de arestas que cada vértice está relacionado (HANSEN, SHNEIDERMAN e SMITH, 2010).

Complementares às regras de agregação em relação a análise quantitativa das redes sociais, as métricas guiadas a vértices buscam identificar a posição de indivíduos em uma rede. Castro (2005) aponta a relevância das medidas de centralidade dos vértices para o estudo das redes sociais:

O objetivo principal é encontrar o papel que cada um desempenha na manutenção e expansão da estrutura da rede [...], permitindo o entendimento da participação dos atores na rede e seu grau de influência e importância. O poder de um ator depende do grau com que ele monopoliza o fluxo de

informação, favores e serviços para e entre os membros de uma rede. [...] O poder é uma propriedade fundamental das redes sociais, ele é uma consequência dos padrões de relacionamentos e de posicionamento na rede. Atores com mais importância ou mais proeminentes usualmente estão em posições estratégicas da rede. Atores inseridos em redes sociais de alta densidade e alto acoplamento têm maior potencialidade para exercer poder (CASTRO, 2005, p.47).

A medida de centralização mais simples é o *grau de centralidade (degree centrality)*, que consiste no número total de conexões ligadas a um vértice. Para redes orientadas, existem duas medidas de grau: *grau de entrada (in-degree)* e *grau de saída (out-degree)*. Grau de entrada diz respeito ao número de conexões que apontam para o vértice, ao passo que grau de saída se refere ao número de ligações que partem do vértice (HANSEN, SHNEIDERMAN e SMITH, 2010).

Geralmente, o grau de centralidade é apresentado na forma normalizada, dividindo-se o número conexões ligadas ao vértice pelo número de ligações em potencial, conforme apresentado na Equação 6. Assim, quando o grau de centralidade for 1, significa que o nó está conectado a todos os outros nós; se zero, significa que o nó não está conectado a nenhum outro nó na rede (CAMPOS, 2014):

$$C_D(i) = \frac{k_i}{N - 1} \quad (6)$$

Onde  $k_i$  representa o número de graus do nó  $i$  e  $N - 1$  representa o número total de nós na rede, com exceção do nó  $i$ .

Castro (2005) explica que em uma rede não-orientada, quanto maior a centralidade de grau, maior a autonomia do nó em relação aos outros nós e, conseqüentemente, maior o poder que esse nó exerce na rede. Por outro lado, Campos (2014, p. 72) mostra que “nós com grau zero estão completamente isolados e, portanto, removê-los da rede não causará impacto nos links presentes”.

Em redes orientadas, um grau de entrada elevado reflete prestígio e popularidade, enquanto que um alto grau de saída representa colaboração e influência (CASTRO, 2005). Entretanto, Hansen, Shneiderman e Smith (2010) lembram que, apesar de ser utilizada algumas vezes como medida de popularidade e prestígio, a centralidade de grau não é recomendada para esse objetivo, uma vez que mede a quantidade de ligações, mas não a qualidade.

A *centralidade de intermediação (betweenness centrality)* é outra métrica bastante difundida na análise de redes sociais. Para Hansen, Shneiderman e Smith (2010, p. 40, tradução nossa), “pode ser compreendida como um tipo de ‘pontuação de ponte’, ou uma medida de o quanto a remoção de um nó da rede vai romper as conexões entre outros nós da rede”. Castro (2005) explica:

A centralidade de intermediação está baseada no controle exercido por um ator sobre as interações entre dois outros atores. Desde que dois atores não sejam adjacentes ou vizinhos, eles dependem de outros atores do grupo para realizar suas trocas. Uma característica negativa é que caso ocorra falha na intermediação, o fluxo entre os atores não adjacentes será interrompido. A posição de intermediação determina influência através da possibilidade de controle do fluxo e indica uma situação preferencial de um ator, pois a intermediação permite a combinação de diferentes ideias que podem influir para a criação de soluções inovadoras. Atores com alta intermediação possuem uma boa visibilidade do que está acontecendo na rede (CASTRO, 2005, p. 49).

Formalmente, a centralidade de intermediação é dada por (WASSERMAN e FAUST, 1994):

$$C_B(n_i) = \frac{\sum_{j < k} g_{jk}(n_i)}{g_{jk}} \quad (7)$$

Onde  $g_{jk}(n)$  representa o número de caminhos mais curtos (caminho geodésico) entre os nós  $j$  e  $k$  que passam por  $n_i$ , enquanto que o denominador  $g_{jk}$  representa o número de caminhos mais curtos entre  $j$  e  $k$ .

Normalizando-se o valor da centralidade de intermediação, temos (WASSERMAN e FAUST, 1994):

$$C'_B(n_i) = \frac{C_B(n_i)}{\frac{(N-1)(N-2)}{2}} \quad (8)$$

A *centralidade de proximidade (closeness centrality)*, em uma perspectiva diferente em relação as outras duas métricas de centralidade apresentadas, captura a distância média entre um vértice e cada um dos demais vértices da rede (HANSEN, SHNEIDERMAN e SMITH, 2010), refletindo o quão integrado ou isolado está um nó da rede (MULLER-PROTHMANN, 2007).

Campos (2014) aponta que a centralidade de proximidade de um nó é inversamente proporcional à distância média dele para qualquer outro nó da rede, gerando um maior potencial de difusão da informação. Castro (2005), mais uma vez, esclarece:

A centralidade de proximidade de um ator se mede pelo número mínimo de passos que ele deve executar para entrar em contato com os outros atores da rede. Quanto mais central for um ator, mais ele está próximo dos outros, e mais rapidamente ele entra em contato ou interage com os outros. É uma medida da autonomia, da independência em relação ao controle exercido pelos outros. Esta medida leva em consideração o somatório dos menores caminhos de um ator para com todos os outros (CASTRO, 2005, p. 48-49).

A centralidade de proximidade pode ser calculada por meio da seguinte equação (WASSERMAN e FAUST, 1994), que representa o somatório das distâncias entre o vértice  $i$  e cada vértice da rede. Como se pretende estimar o quão cada vértice  $i$  está próximo dos demais vértices (e não o quão distante), o somatório é elevado à potência -1 (CAMPOS, 2014):

$$C_c(n_i) = \left[ \sum_{j=1}^g d(n_i, n_j) \right]^{-1} \quad (9)$$

Para normalização da métrica de centralidade de proximidade, multiplica-se o valor encontrado na equação anterior pelo número de vértices, menos um. Assim, conforme apresentam Wassermann e Faust (1994) obtêm-se:

$$C'_c(n_i) = C_c(n_i) \cdot (g - 1) \quad (10)$$

Assim, quanto mais próximo de um for a medida de centralidade de proximidade de um vértice, menor é sua distância média em relação aos demais vértices da rede; conforme o valor da centralidade de proximidade de aproxima de zero, mais distante está o vértice em relação ao restante da rede (CAMPOS, 2014).

Outra métrica bastante utilizada na análise de redes sociais é o PageRank. O algoritmo para seu cálculo foi apresentado pelos criadores do Google, Sergey Brian e Larry Page, em 1998 (PAGE *et. al.*, 1999). Brian e Page identificaram que os links que conectam as páginas web carregam uma importante informação: eles indicam um implícito transporte de autoridade para as páginas que recebem esse apontamento.

Nesse sentido, o algoritmo interpreta o hiperlink de uma página  $x$  para uma página  $y$  como um voto, dado de  $x$  para  $y$ , ao mesmo tempo em que analisa a página que lança o voto: votos dados por páginas mais importantes têm um maior peso e ajudam a melhorar a importância das páginas que o recebem (LIU e YU, 2008).

Assim, a importância ou prestígio de uma página  $i$ , representada por  $P(i)$ , pode ser determinada pela soma da pontuação de PageRank de todas as páginas que apontam para  $i$ ; além disso, a pontuação de prestígio de uma página  $j$  deve ser compartilhada por todas as páginas que recebem um voto de  $j$  (LIU e YU, 2008).

Esse mesmo conceito pode ser estendido a qualquer rede direcionada. Deste modo, considerando a rede social como um grafo direcionado  $G = (V, E)$ , no qual o conjunto de vértices  $V$  são os nós da rede e o conjunto de arestas  $E$  representam as ligações entre esses vértices, o PageRank pode ser definido por:

$$P(i) = \sum_{(j,i) \in E} \frac{P(j)}{O_j} \quad (11)$$

Onde  $O_j$  é o grau de saída do vértice  $j$ .

Matematicamente, temos um sistema de  $n$  equações lineares para calcular os  $n$  PageRank dos  $n$  vértices da rede, que podem ser representadas por meio de uma matriz  $n$ -dimensional  $P$ , dada por:

$$P = ((P(1), P(2), \dots, P(n))^T \quad (12)$$

Sendo  $A$  a matriz de adjacência do grafo  $G$ :

$$A_{ij} = \begin{cases} 1/O_i, & \text{se } (i, j) \in E \\ 0, & \text{caso contrário} \end{cases} \quad (13)$$

Pode-se escrever o sistema de  $n$  equações como:

$$P = A^T P \quad (14)$$

Uma vez que esta é uma definição circular, um algoritmo iterativo deve ser utilizado para resolvê-la, por meio de uma técnica denominada *power iteration* (GOLUB e VAN LOAN, 2012). Para satisfazer as condições dessa técnica, um parâmetro de amortização  $d$  é utilizado, levando a equação para o cálculo do PageRank de cada vértice da rede, dada por:

$$P(i) = (1 - d) + d \sum_{j=1}^n A_{ji} P(j) \quad (15)$$

Na equação anterior, o valor de  $P$  pode ser configurado entre 0 e 1, sendo o valor 0,85 mais comumente utilizado (LIU e YU, 2008). A matriz do PageRank para a primeira iteração é uma matriz identidade. A equação acima é então aplicada por  $m$  iterações, até que os valores para  $P(i)$  sejam estabilizados.

#### 2.2.6 Redes sociais virtuais

As redes sociais virtuais – por vezes referenciadas por *sites* de redes sociais ou redes sociais *online* – podem ser caracterizadas como um serviço disponibilizado na internet que possibilita que seus usuários construam perfis públicos ou semi-públicos dentro de um sistema, articulem uma lista de outros usuários com os quais compartilha conexões e, por fim, visualizem e percorram sua lista de conexões assim como outras listas criadas por usuários do sistema (ELLISSON et al, 2007; BENEVENUTO, ALMEIDA e SILVA, 2011). Com base na definição apresentada, Benevenuto, Almeida e Silva (2011) relacionam as principais redes sociais com seus respectivos endereços de acesso, classificando-as quanto ao propósito.

**QUADRO 1 – PERSPECTIVA HISTÓRICA DOS PRINCIPAIS SITES DE REDES SOCIAIS, APRESENTANDO SEUS PROPÓSITOS, ENDEREÇOS E NÚMERO DE USUÁRIOS**

Nome	Propósito	Endereço (URL)	Usuários <sup>1</sup>
SixDegrees.com (1997)	Amizades	<a href="http://sixdegrees.com">http://sixdegrees.com</a> <sup>2</sup>	-
LiveJournal (1999)	Blogs, jornais e diários	<a href="http://www.livejournal.com">http://www.livejournal.com</a>	-
LinkedIn (2003)	Profissionais	<a href="https://www.linkedin.com">https://www.linkedin.com</a>	187 Mi
MySpace (2003)	Amizades	<a href="https://myspace.com">https://myspace.com</a>	-
LastFM (2003)	Compartilhamento de rádio/músicas	<a href="http://www.lastfm.com">http://www.lastfm.com</a>	-
Hi5 (2003)	Amizades	<a href="http://www.hi5.com">http://www.hi5.com</a>	-
Orkut (2004)	Amizades	<a href="https://orkut.google.com">https://orkut.google.com</a> <sup>3</sup>	-
Flickr (2004)	Compartilhamento de fotos	<a href="https://www.flickr.com">https://www.flickr.com</a>	-
Digg (2004)	Compartilhamento de <i>bookmarks</i>	<a href="http://digg.com">http://digg.com</a>	-
YouTube (2005)	Compartilhamento de vídeos	<a href="http://www.youtube.com">http://www.youtube.com</a>	1,0 Bi
Twitter (2006)	Troca de mensagens curtas	<a href="https://twitter.com">https://twitter.com</a>	288 Mi
Facebook (2006)	Amizades	<a href="https://www.facebook.com">https://www.facebook.com</a>	1,39 Bi
Instagram (2010)	Compartilhamento de fotos	<a href="https://plus.google.com">https://plus.google.com</a>	300 Mi
Pinterest (2010)	Compartilhamento de fotos	<a href="https://www.pinterest.com">https://www.pinterest.com</a>	40 Mi
Google Plus (2011)	Amizades	<a href="https://plus.google.com">https://plus.google.com</a>	540 Mi

FONTE: ADAPTADO DE ELLISSON et al (2007); BENEVENUTO, ALMEIDA e SILVA (2011); ALLTON (2015)

Como se pode perceber a partir do Quadro 1, o primeiro *site* de rede social foi o SixDegrees.com, lançado no ano de 1997. Apesar de não ser o primeiro *site* a permitir a criação de perfis públicos, a manutenção de uma rede de amizade e a navegação nessa lista de amigos, foi pioneiro em reunir todas essas características em uma única plataforma, tornando-se uma ferramenta que auxiliava a comunicação e troca de mensagens entre amigos. Apesar do sucesso inicial, a plataforma sucumbiu em 2000, quando se tornou um negócio economicamente inviável (ELLISSON et al, 2007).

A partir de 2003, observou-se um “boom” no lançamento de redes sociais, que passaram a explorar vários outros nichos, sobretudo profissionais e relacionadas ao compartilhamento de conteúdo digital, como fotos, músicas e vídeos. A expansão dos sites de redes sociais observada na primeira metade da década de 2000 culminou no surgimento de duas das principais redes sociais que conhecemos hoje, ambas

<sup>1</sup> Número de usuários mensalmente ativos

<sup>2</sup> Site de rede social desativado em 2000

<sup>3</sup> Site de rede social desativado em 2014

lançadas para o público em geral em 2006: Twitter e Facebook. Após esse período, outras importantes redes sociais foram lançadas, como o Instagram, Pinterest e Google Plus.

Comparados com os sites de redes sociais tradicionais, os microblogs, como o Twitter, preenchem a necessidade de uma rápida forma de comunicação, quase que instantânea, viabilizada pelo emprego de mensagens curtas, possibilitando aos usuários um menor investimento de tempo e esforço para geração de conteúdo. Essa particularidade dos microblogs faz com que as atualizações sejam mais frequentes, em comparação com mídias sociais que não possuem limitação em relação ao tamanho das mensagens (JAVA et. al., 2007). Essa mídia social é mais bem descrita na seção 2.2.6.1.

#### 2.2.6.1 *Twitter*

O Twitter é uma rede social virtual em formato de microblog lançada em 2006 por Jack Dorsey, Biz Stone e Evan Williams, na qual os usuários podem publicar mensagens de texto de até 140 caracteres, conhecidas como *tweets*. Essa limitação de caracteres se dá pelo fato de que inicialmente os *tweets* eram enviados exclusivamente via SMS, o que facilitava a atualização em tempo real da linha do tempo de cada usuário. Apesar da evolução da tecnologia móvel ter permitido o envio de mensagens por outros formatos além do SMS, essa mídia social manteve suas mensagens curtas como característica. Atualmente, além do formato original, as atualizações podem ser postadas e lidas a partir de navegadores *web* ou ainda por aplicações desenvolvidas por terceiros para as mais diversas plataformas, como *desktops*, *smartphones* e outros dispositivos móveis (MARWICK et. al., 2011; LIMA, 2012).

Ao contrário da maioria de outros sites de redes sociais, como o Facebook, por exemplo, as relações sociais construídas no Twitter não requerem reciprocidade, constituindo, portanto, um modelo direcionado de ligação. Assim, um usuário pode seguir (*follow*) qualquer perfil de seu interesse, passando a receber todas as atualizações (*tweets*) do perfil seguido (*followed*); entretanto, o perfil seguido não precisa segui-lo de volta, necessariamente (KWAK et. al., 2010; MARWICK et. al., 2011).

Apesar de a maior parte das postagens no Twitter não ser direcionada – portanto, destinada a todos os usuários que seguem o perfil responsável pela publicação – esse site de rede social possibilita ainda que as mensagens sejam dirigidas, convencionadas pela inserção, no início da postagem, do símbolo “@” (arroba) seguido pelo nome do usuário para o qual o *tweet* se destina (HUBERMANN, ROMERO e WU, 2008). Essa prática recebe também o nome de endereçamento (BOYD, GOLDER e LOTAN, 2010).

Ainda que a mensagem seja direcionada para um usuário específico, é importante ressaltar que ela continua sendo pública, podendo ser visualizada pelos demais usuários. As postagens direcionadas correspondem a cerca de 25,4% das mensagens dessa mídia social, o que mostra que esse é um recurso bastante utilizado pelos usuários da rede (HUBERMANN, ROMERO e WU, 2008). O símbolo “@” seguido pelo nome de um usuário, quando não no início da postagem, também é utilizado para referenciar outros usuários em mensagens não-direcionadas. (BOYD, GOLDER e LOTAN, 2010).

Outra convenção de sintaxe é a demarcação de tópicos da postagem por meio da combinação do símbolo “#” (*hashtag*) seguido por uma palavra-chave. Essa prática de usar palavras-chave para categorizar os *tweets* é uma herança do uso de rótulos de agrupamento amplamente empregados para classificar conteúdo web, também conhecidos como *tags* (BOYD, GOLDER e LOTAN, 2010).

Também é uma prática bastante comum entre os usuários do Twitter a utilização do símbolo “RT” (*retweet*) para propagar a postagem de outro usuário em sua rede, o que pode ser realizado por meio da cópia do *tweet* original precedido pelo símbolo RT e pelo endereçamento do autor de origem. Assim, o *retweet* da postagem “Olá Mundo!” realizada originalmente pelo autor “A”, teria o seguinte formato: “RT @A Olá Mundo!”. A análise dos *retweets* é uma importante ferramenta para compreender como as mensagens são disseminadas pela rede (BOYD, GOLDER e LOTAN, 2010).

### 2.3 MINERAÇÃO DE OPINIÃO

Nessa seção, são apresentados os principais fundamentos relacionados à mineração de opinião, partindo desde a definição de dado, informação e conhecimento, que é exposta na seção 2.3.1. O KDD (*Knowledge Discovery in Databases*), processo de descoberta de conhecimento em base de dados, é

apresentado na seção 2.3.2, sendo que sua principal etapa, a mineração de dados, é detalhada na seção 2.3.3. Mais adiante, na seção 2.3.4, são discutidas as aplicações da mineração de opinião nas redes sociais e como suas técnicas podem auxiliar na descoberta de conhecimento com base nos dados disponíveis nessas mídias.

### 2.3.1 Dado, informação e conhecimento

As fronteiras que delimitam os conceitos de dados, informação e conhecimento são bastante sutis, sendo esse um dos motivos de esses termos serem utilizados incorretamente até mesmo por profissionais da área de TI ou gestores de grandes organizações.

Davenport e Prusak (1998) definem dados como um conjunto de fatos distintos e objetivos, relativos a eventos e descritos ou ainda registros estruturados de transações. Por si só, os dados têm pouca relevância ou propósito, sendo apenas uma observação do estado do mundo. E é nesse ponto que está a sutil diferença entre dado e informação.

Segundo Druker (2006), poderíamos definir a informação como sendo dados que carregam consigo um significado, tendo certa relevância e algum propósito. Corroborando com essa afirmação, Angeloni (2003) considera a informação como sendo dados processados e contextualizados. Malhotra (1993), define a informação como a matéria-prima utilizada como insumo para a obtenção do conhecimento.

Seguindo o aumento da escala de valor agregado, chegamos a definição de conhecimento, que pode ser conjecturado como uma informação valiosa a mente humana, que é gerado a partir de uma reflexão, requer uma capacidade de síntese e deve estar contextualizado, além de ser frequentemente tácito, ou seja, esta geralmente implícito na mente dos conhecedores (DAVENPORT, 2000).

Angeloni (2003), a partir da análise da relação entre dado, informação e conhecimento, infere que “os dados por si só não significam conhecimento útil para a tomada de decisão, constituindo apenas o início do processo”. Ainda segundo a autora, os responsáveis pela tomada de decisão dentro das organizações têm como grande desafio a transformação dos dados em informação e desta em conhecimento.

A aplicação do KDD sobre uma base de dados para descoberta de conhecimento pode ser importante catalisador nessa transformação. O detalhamento desse processo é feito na Seção 2.3.2.

### 2.3.2 Descoberta de conhecimento

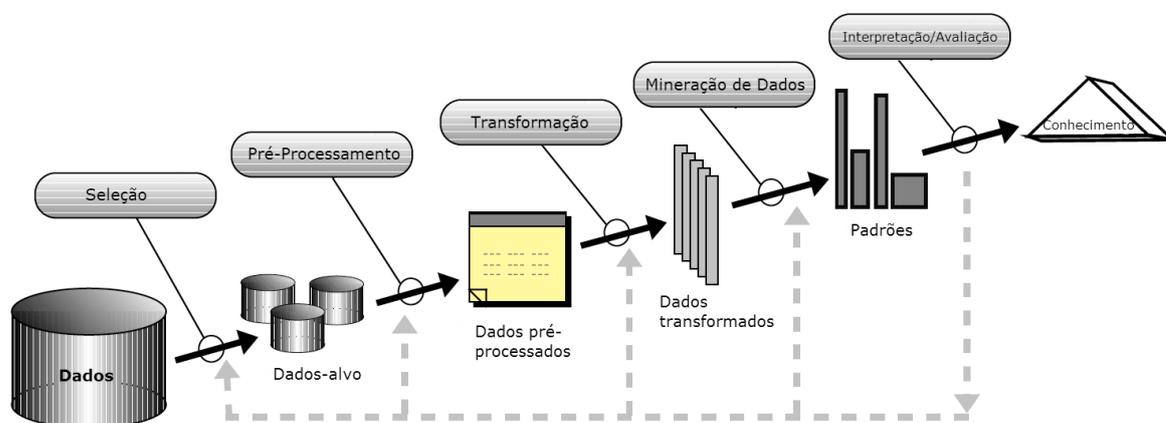
Uma das técnicas que tem se mostrado mais eficientes para análise de grande volume de dados é conhecida como *Knowledge Discovery in Databases* ou Descoberta de Conhecimento em Bases de Dados.

Esse termo foi utilizado pela primeira vez em um workshop que aconteceu no dia 20 de agosto de 1989, em Detroit, nos EUA, quando se identificou a oportunidade e necessidade de se extrair conhecimento a partir de bases de dados, uma vez que, já naquela época, a quantidade de registros encontrados nas bases de dados não implicava necessariamente em informação útil, ou seja, em conhecimento (PIATETSKY-SHAPIRO, 1990, p. 68).

Mais tarde, Frawley, Piatetsky-Shapiro e Matheus (1992, p. 57) definiram o termo como sendo o processo não trivial de extração de informação implícita, previamente desconhecida e potencialmente útil, a partir dos dados. Em outras palavras, o que o KDD propõe é exatamente a transformação de dados (registros que sozinhos não possuem nenhum significado e que geralmente são bastante volumosos e de difícil compreensão) em algo útil e compreensivo (conhecimento).

Fayyad, Piatetsky-Shapiro e Smyth (1996, p.37) definem cinco atividades principais que formam o processo KDD (Figura 19): seleção, pré-processamento, transformação, mineração de dados e avaliação.

**FIGURA 19 – VISÃO GERAL DAS ATIVIDADES QUE CONSTITUEM O PROCESSO DE DESCOBERTA DE CONHECIMENTO (KDD)**



FONTE: ADAPTADO DE FAYYAD, PIATETSKY-SHAPIRO e SMYTH (1996, p. 37)

A seleção consiste no entendimento do domínio da aplicação com o objetivo de identificar um subconjunto de dados a partir do qual se pretende extrair alguma informação; é importante que nessa etapa o objetivo a ser alcançado com a aplicação do KDD esteja definido.

A etapa de pré-processamento inclui a execução de tarefas para a remoção de ruídos (*outliers*) e definição de estratégias para tratamento de registros com campos que possuem valores ausentes.

A transformação, por sua vez, tem por objetivo a redução (eliminação) ou estratificação (divisão em classes) de alguns atributos. A transformação deve levar em consideração o objetivo definido no início do processo.

A fase de mineração de dados consiste na aplicação de algoritmos de mineração de dados sobre os dados transformados com o objetivo de identificar padrões ou modelos.

Por fim, a fase de avaliação busca interpretar e avaliar os padrões obtidos na fase de mineração de dados. Se estes padrões descobertos forem novos, úteis e compreensíveis (decisão que exige o envolvimento de um analista humano neste processo), o KDD atinge seu objetivo maior que é a geração de conhecimento.

### 2.3.3 Mineração de dados

Ao passo que KDD é o processo abrangente de descoberta de conhecimento a partir de dados, a etapa de mineração de dados consiste na aplicação de algoritmos específicos para a extração de padrões a partir destes dados, geralmente empregando aprendizagem de máquina, reconhecimento de padrões e técnicas relacionadas à estatística (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996, p. 37).

Corroborando com essa distinção, Lobur *et. al.* (2008, p. 95) ressaltam mais uma vez que mineração de dados é um passo dentro do processo KDD, “consistindo na aplicação de análise de dados e algoritmos de descoberta que, sob limitações aceitáveis de eficiência computacional, produzem um particular número de padrões a partir dos dados”.

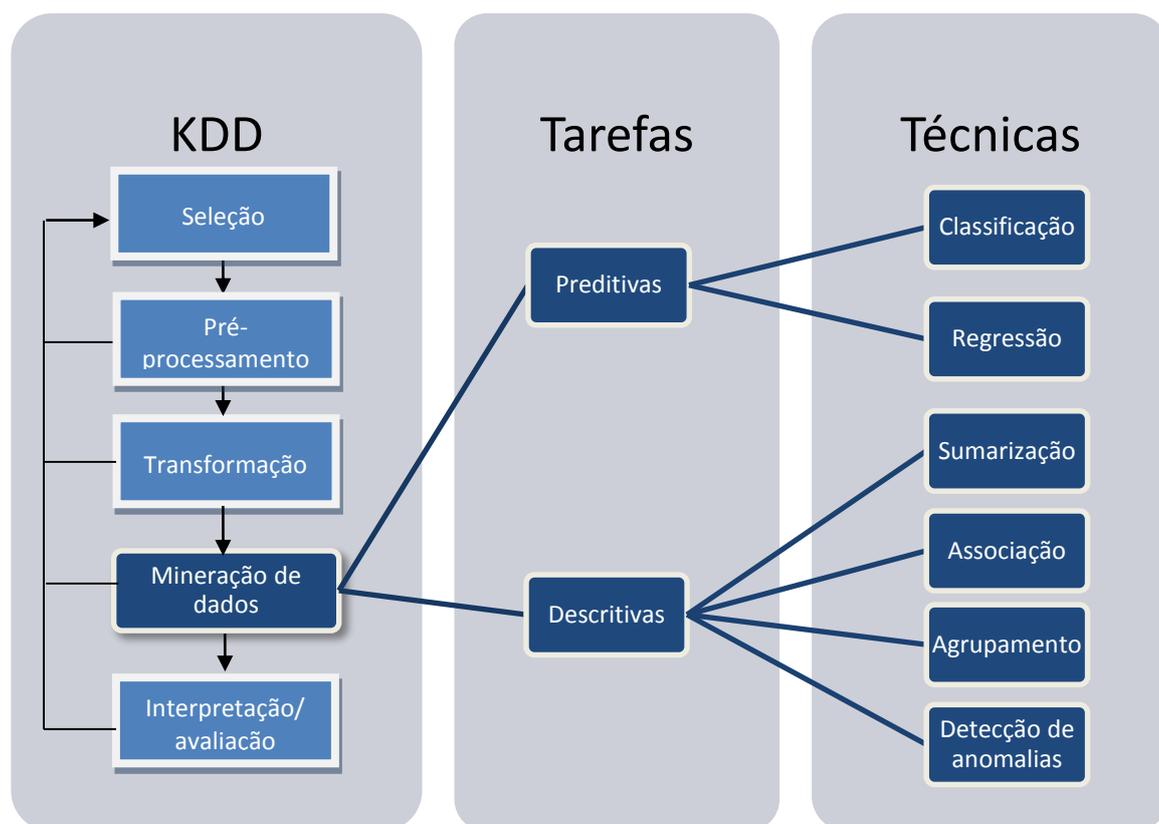
As tarefas de mineração de dados podem ser preditivas ou descritivas com base no tipo de conhecimento que se deseja extrair a partir do conjunto de dados (HAN e KAMBER, 2006). Enquanto as tarefas preditivas buscam descobrir o valor de atributos (denominados variáveis-meta, variáveis dependentes ou variáveis-alvo) com

base em outras propriedades do objeto de estudo (conhecidas por variáveis independentes ou explicativas), as tarefas descritivas têm como objetivo resumir os relacionamentos presentes no conjunto de dados por meio da identificação de padrões (TAN, STEINBACH e KUMAR, 2009).

As tarefas preditivas podem ser implementadas por meio de técnicas de classificação ou regressão, ao passo que as tarefas descritivas contemplam as técnicas de sumarização, associação, agrupamento e ainda detecção de anomalias (FAYYAD, PIATETSKY-SHAPIO e SMYTH, 1996; LAROSE, 2005; SUMATHI e SIVANANDAM, 2006; OLSON e DELEN, 2008; TAN, STEINBACH e KUMAR, 2009).

A Figura 20 destaca a mineração de dados como um importante passo dentro do processo de descoberta de conhecimento, mostrando suas tarefas e as técnicas mais comuns por meio das quais as tarefas podem ser implementadas.

**FIGURA 20 – A MINERAÇÃO DE DADOS COMO UMA IMPORTANTE ETAPA DO PROCESSO KDD E PRINCIPAIS TAREFAS E TÉCNICAS RELACIONADAS**



FONTE: O AUTOR (2015)

Na técnica de classificação, um modelo ou função é construído a partir da análise da relação entre os atributos e as classes dos objetos de um conjunto de

treinamento; esse modelo é então utilizado para classificar novos objetos com bases em seus atributos (SUMATHI e SIVANANDAM, 2006; OLSEN e DELEN, 2008). As técnicas de classificação são usualmente implementadas por meio de métodos de aprendizado supervisionado como árvores de decisão e classificação bayesiana (HAN e KAMBER, 2006). São exemplos de algoritmos de classificação: ID3 (QUINLAN, 1986), C4.5 (QUINLAN, 1993) – ambos geradores de árvores de decisão – Naive Bayes (NB) (JOHN e LANGLEY, 1995) e *Support Vector Machines* (SVM) (VAPNIK, 2006). Os algoritmos C4.5 e SVM, aplicados nessa pesquisa, são detalhados nas Subseções 2.3.3.1 e 2.3.3.2, respectivamente.

Enquanto que as técnicas de classificação buscam encontrar modelos de predição para determinação de variáveis-meta discretas (classes), técnicas de regressão têm como objetivo encontrar funções para descoberta de variáveis-meta do tipo contínuas, com base nos demais atributos do objeto (LAROSE, 2005; TAN, STEINBACH e KUMAR, 2009). A predição faz uso de metodologias estatísticas como, por exemplo, a análise de regressão (HAN e KAMBER, 2006).

A sumarização nada mais é do que a abstração ou generalização dos dados. Nas técnicas de sumarização, busca-se extrair informações condensadas a partir da massa de dados, resultando em um conjunto menor que fornece uma visão mais geral dos dados; tal visão pode potencialmente consistir em informação útil e relevante para tomada de decisão (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996; SUMATHI e SIVANANDAM, 2006).

As técnicas de associação têm como objetivo encontrar relações ou dependências entre atributos de um objeto a partir da análise de conjuntos de dados; essas dependências são geralmente expressas por meio de regras, que indicam relações de causa e efeito entre atributos de um objeto (SYMEODINIS e MITKAS, 2006; CIOS et. al., 2007). Técnicas de associação são implementadas por meio do método de aprendizado não supervisionado conhecido como regras de associação (CIOS et. al., 2007). O algoritmo mais utilizado é o Apriori (AGRAWAL e SRIKANT, 1994; LIU, HSU e MA, 1998).

Técnicas de agrupamento buscam identificar grupos de objetos a partir dos seus atributos. Nesse contexto, os grupos também podem ser conhecidos como *clusters*. Esses grupos são formados de modo a maximizar a similaridade entre objetos dentro de um mesmo grupo e minimizá-la entre objetos de grupos diferentes (LAROSE, 2005; SUMATHI e SIVANANDAM, 2006).

Nas técnicas de agrupamento, os rótulos das classes não estão presentes no conjunto de treinamento porque geralmente não são conhecidas inicialmente. Assim, faz-se uso dessa técnica para gerar classes a serem utilizadas como entrada para técnicas de classificação (HAN e KAMBER, 2006). Técnicas de agrupamento são implementadas por meio de método de aprendizado não supervisionado chamado análise de agrupamento. O algoritmo mais comum que implementa esse método é o K-means (HARTIGAN, 1979).

A *detecção de anomalias*, também conhecida como detecção de *outliers* consiste em uma técnica de mineração de dados bastante utilizada nas etapas de pré-processamento do KDD para identificação e remoção de registros discrepantes que não se enquadram no comportamento dos demais registros do conjunto de dados. Geralmente, esses registros atípicos são ruídos dentro do conjunto de dados e, se não removidos, podem impactar negativamente na qualidade dos resultados encontrados (HAN e KAMBER, 2006; SYMEONIDIS e MITKAS, 2006). A detecção de anomalias pode ser realizada por meio de métodos baseados em classificação, agrupamento, baseados em distância e estatísticos (CHANDOLA, BANERJEE e KUMAR, 2009). Os algoritmos mais comuns que implementam métodos de detecção de anomalias são o Orca, Load e Reloaded (OTEY, PARTHASARATHY e GHOTING, 2005).

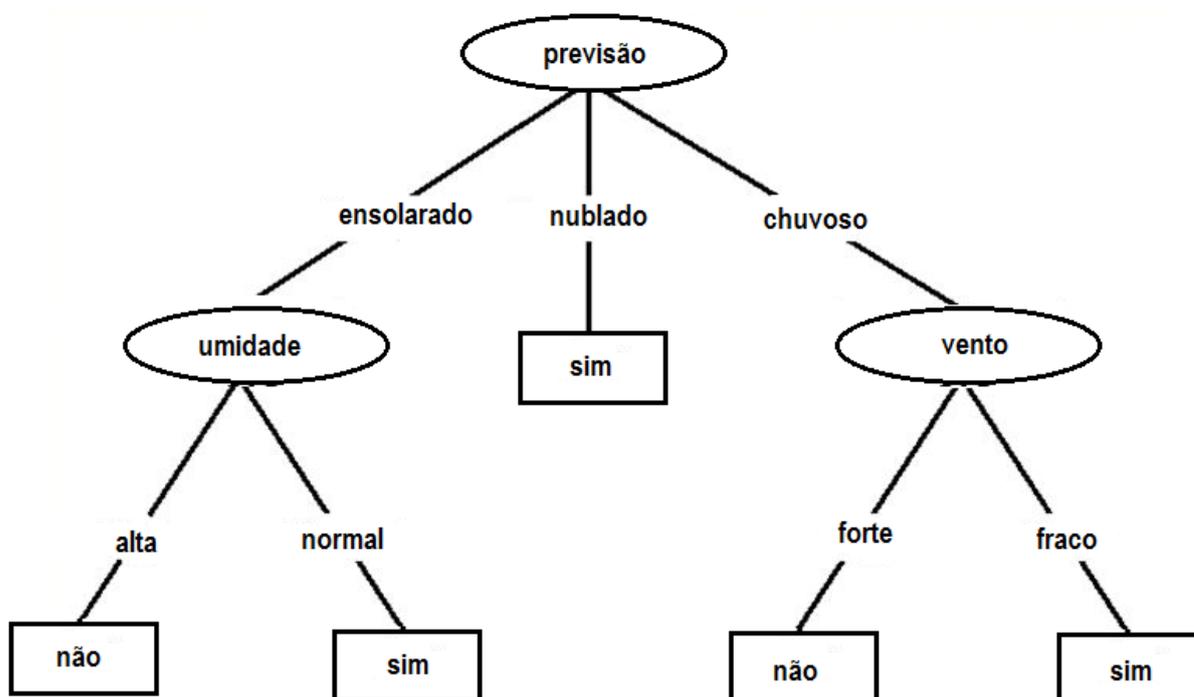
#### 2.3.3.1 C4.5

O C4.5 é um algoritmo de classificação cuja saída consiste em uma árvore de decisão formada por folhas – que representam as classes – e nós de decisão – que especificam um teste a ser realizado sobre um atributo preditor; cada nó de decisão possui um galho ou sub-árvore para cada possível resultado do teste (QUINLAN, 1993).

Partindo do nó de decisão inicial, também conhecido como nó raiz – que por convenção é colocado no topo da árvore de decisão – os atributos preditores são testados. O resultado obtido em cada teste leva ao próximo nó de decisão. Essa operação é repetida até que um nó folha seja alcançado (LAROSE, 2005).

Um exemplo clássico de árvore de decisão para inferência sobre a ocorrência de um jogo de tênis a partir das condições climáticas é mostrado na Figura 21.

FIGURA 21 – ÁRVORE DE DECISÃO PARA INFERÊNCIA SOBRE OCORRÊNCIA DE UM JOGO DE TÊNIS



FONTE: ADAPTADO DE MITCHELL (1997)

Na Figura 21, podem ser observadas duas classes (sim e não) e três atributos preditores (previsão, umidade e vento). O atributo previsão pode assumir os valores ensolarado, nublado ou chuvoso, enquanto que umidade pode ser alta ou normal e o vento pode ser forte ou fraco. Assim, como base nessa árvore de classificação pode-se dizer, por exemplo, que haveria jogo caso o tempo estivesse ensolarado e a umidade normal; por outro lado, não haveria uma partida de tênis se, por exemplo, o tempo estivesse chuvoso e com ventos fortes.

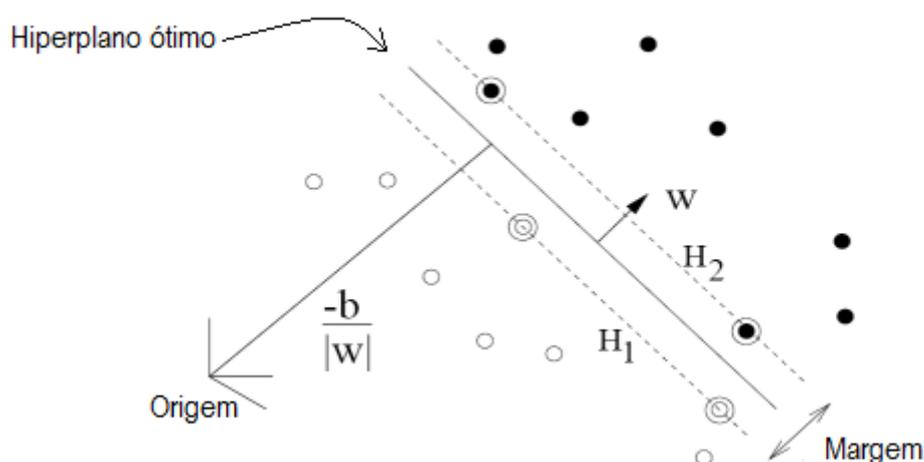
### 2.3.3.2 SVM

O algoritmo SVM foi originalmente proposto pela matemático Vladimir Vapnik em 1979, que o empregou a técnica para classificação de dados em duas classes (VAPNIK, 2006). Conforme afirma Oguri (2006), essa técnica consiste essencialmente em uma abordagem geométrica para o problema de classificação, onde cada entrada do conjunto de treinamento pode ser visto como um ponto  $x_i$  em um espaço  $\mathbb{R}^M$ , de modo que o aprendizado consiste em dividir os elementos nesse

espaço euclidiano, separando, no caso de classificação de textos, os pontos que representam os textos positivos daqueles que representam os textos negativos.

Em sua implementação mais simples, conhecida como SVM linear, assume-se que as amostras de treinamento das duas classes são linearmente separáveis. A partir dessa premissa, o algoritmo SVM busca encontrar uma função de decisão em que a distância entre os conjuntos das amostras de treinamento é maximizada. Essa função que maximiza a separação entre os dois conjuntos é conhecida como função ótima; de modo análogo, o hiperplano originado a partir dessa função é denominado hiperplano ótimo (ANDREOLA, 2009). O hiperplano ótimo e a margem maximizada de separação são mostrados na Figura 22.

**FIGURA 22 – SOLUÇÃO BASEADA NO SVM PARA DUAS CLASSES LINEARMENTE SEPARÁVEIS**



**FONTE: OGURI (2006)**

O hiperplano pode ser definido pela equação de primeiro grau  $x \cdot w + b = 0$  onde  $x$  são os pontos sobre o hiperplano,  $w$  é a normal ao hiperplano e  $(-b) / (|w|)$  é a distância do hiperplano à origem. Os pontos sobre os quais a linha pontilhada passa formam os vetores de suporte (OGURI, 2006).

#### 2.3.4 Mineração de opinião: definição, tarefas e abordagens

Dada a grande expansão das mídias sociais por meio da Web, organizações e indivíduos tem utilizado seu conteúdo para tomada de decisão. Contudo, ao contrário

das bases tradicionais de mineração de dados, os registros encontrados na Web estão na forma de textos não estruturados.

Para tentar resolver esses problemas, sistemas automatizados de análise de sentimentos são necessários (LIU, 2007). A análise automática de sentimentos consiste na "extração automática de opiniões a partir de uma fonte de dados não estruturada como textos, áudio e vídeo" (BOIY e MOENS, 2009, p. 1, tradução nossa). Liu (2010, p. 629, tradução nossa) delimita a definição proposta por Boiy e Moens (2009): "a análise de sentimentos ou mineração de opinião é o estudo computacional de opiniões, sentimentos e emoções expressas em texto".

Por sua vez, Chen e Zimbra (2010) tratam a mineração de opinião como uma subdisciplina de mineração de dados, congregando um conjunto de técnicas para extração, classificação, compreensão e avaliação de opiniões expressas por meio de comentários realizadas por usuários em sites de notícias, mídias sociais ou qualquer outro conteúdo *online* gerado por usuários que expressam sua opinião sobre determinado assunto, indivíduo, organização, produto ou serviço. Para isso, reúne esforços das áreas de processamento de linguagem natural, recuperação de informação e inteligência artificial (BECKER e TUMITAN, 2013).

Vinodhini e Chandrasekaran (2012) definem análise de sentimentos por meio da relação entre um público e um produto, destacando-a como um importante ferramental para a análise das mídias sociais:

Análise de sentimentos é um tipo de processamento de linguagem natural para acompanhar o estado de espírito de um público a respeito de um produto ou tópico. A análise de sentimentos, que também é conhecida como mineração de opinião, envolve a construção de um sistema para coletar e analisar opiniões sobre produtos feitas através de postagens em blogs, comentários, *reviews* ou *tweets* (VINODHINI e CHANDRASEKARAN, 2012, p. 282, tradução nossa).

O insumo utilizado para essa análise são os chamados "textos avaliativos", que nada mais são do que documentos por meio dos quais as opiniões são expressas (LIU, 2007). O autor, em concordância com a definição apresentada por Vinodini e Chandrasekaran (2012), exemplifica que esses documentos podem ser textos de *review* de um produto em um *web commerce*, postagens em um microblog ou discussões em fóruns especializados.

Na análise de sentimentos, existem algumas palavras que são comumente utilizadas para expressar sentimentos positivos ou negativos: são conhecidas como

"palavras de sentimento" ou "palavras de opinião". Palavras como "bom", "legal", "ótimo" são indicativas de sentimentos positivos, ao passo que "ruim", "desagradável", "péssimo" são palavras que expressam opiniões negativas. Além dessas palavras, existem ainda frases ou expressões idiomáticas que também expressam sentimentos, como é o caso de "custou os olhos da cara". Ao conjunto de palavras e expressões idiomáticas dá-se o nome de "léxico de sentimentos" ou "léxico de opiniões" (LIU, 2012).

Embora o léxico seja uma parte importante da análise de sentimentos, não é suficiente, dada a grande complexidade dessa tarefa. Além das dificuldades da análise de sentimentos herdadas dos problemas inerentes ao processamento de linguagem natural, como gerenciamento de negações e ambiguidade de palavras, existem ainda outros desafios a serem enfrentados (LIU, 2012). O surgimento de novos vocabulários, abreviações de palavras existentes, jargões específicos de nichos profissionais, expressões regionais, erros ortográficos e gírias são apenas alguns deles (BOIY e MOENS, 2009; BECKER e TUMITAN, 2013).

Outra dificuldade consiste no fato de uma palavra poder adquirir conotação positiva ou negativa dependendo do contexto em que é aplicada (LIU, 2012). Um exemplo é o verbo "apavorar". Aplicada na frase "essa banda apavora", apresenta uma conotação positiva. Por outro lado, na frase "aquele restaurante me apavora", demonstra um sentimento negativo.

Além disso, nem sempre uma expressão que contenha uma "palavra de sentimento" expressa, de fato, uma opinião, o que é muito comum em questões interrogativas e frases condicionais (LIU, 2012). Tanto na frase "o produto da marca ACME é bom?" quanto em "se eu encontrar uma camisa legal, eu vou comprar" não se pode dizer que exista um sentimento positivo.

Entretanto, Liu (2012) lembra que nem sempre frases interrogativas e condicionais implicarão em ausência de sentimentos. As sentenças "alguém mais tem esse celular que vive aquecendo?" e "se você quer comprar um bom celular, compre da marca ACME" apresentam, respectivamente, a noção de sentimento negativo e positivo, mesmo sendo frases interrogativas e condicionais.

Outro desafio da análise de sentimentos é a presença de sarcasmo em algumas frases (LIU, 2012), como na expressão "nossa, que governante bom, não sei como poderia viver sem ele". O autor salienta que esse tipo de comportamento não é

tão comum quando se avalia produtos ou serviços, sendo mais recorrente em opiniões políticas.

Existem ainda sentenças que expressam opinião sem necessariamente conter palavras de sentimentos, como em "minha máquina de lavar utiliza muita água". Assim como nesse exemplo, geralmente consiste em opiniões objetivas utilizadas para representar informações factuais, como a má utilização de um recurso (LIU, 2012).

Para tentar superar esses desafios, técnicas de análise de sentimentos vêm sendo desenvolvidas ao longo dos anos para explorar as várias fontes de textos não estruturados disponíveis na *web* para auxiliar empresas e indivíduos a obter de modo mais efetivo possível informação útil (LIU, 2007). Isso pode ser realizado por meio de três principais tarefas de mineração. Embora a mineração de opinião possa ser muitas vezes confundida com a atribuição de polaridade aos textos (PANG e LEE, 2008) – talvez pelo fato de ser o tópico mais amplamente estudado nessa área (LIU, 2010) – existem outras duas tarefas relacionadas a mineração de opinião: mineração de opinião baseada em características; e mineração de relacionamento e sentenças comparativas. Essas tarefas de análise automática de textos avaliativos – que consistem em documentos que expressam opiniões – são apresentadas a seguir, com maior enfoque para classificação de sentimentos, que será utilizada para a análise dos dados dessa pesquisa.

#### 2.3.4.1 *Classificação de Sentimentos*

A atribuição de polaridade aos textos é a tarefa de mineração de opinião mais comum, conhecida formalmente como "classificação de polaridade" (PANG e LEE, 2008; CAMBRIA et. al. 2013), "classificação de sentimentos" (LIU, 2007) ou ainda "classificação de sentimentos em nível de documento" (LIU e ZHANG, 2012) e tem como objetivo classificar um texto avaliativo como positivo ou negativo (ou neutro, em algumas técnicas), preocupando-se em realizar uma rápida determinação da opinião que prevalece em relação ao documento como um todo, sem se importar em descobrir especificamente quais atributos do objeto alvo agradaram ou não o avaliador (LIU, 2007).

Para essa tarefa, assume-se que o documento opinativo (por exemplo, a revisão de um produto ou um *tweet*) revela opiniões a respeito de uma única entidade expressas por um único indivíduo, chamado detentor da opinião (LIU e ZHANG, 2012).

Quatro principais abordagens para a classificação de sentimentos – léxica, baseada em aprendizado de máquina, estatísticas e semânticas – vem sendo utilizadas, geralmente combinadas para obtenção de melhores resultados (TSYTSARAU e PALPANAS, 2012; BECKER e TUMITAN, 2013).

#### 2.3.4.1.1 Abordagem léxica

A abordagem léxica – também conhecida como abordagem baseada em dicionário ou abordagem linguística – faz uso de um léxico de sentimentos, que consiste em um conjunto de palavras e expressões (cada qual com suas possíveis flexões) associadas a sua respectiva polaridade, que pode ser expressa por meio de uma categoria ou por um valor dentro de uma escala contínua (BECKER e TUMITAN, 2013). Nesses casos, a polaridade de uma sentença ou documento é geralmente calculada por meio da média aritmética (simples ou ponderada) das polaridades de cada palavra presente no texto analisado (TSYTSARAU e PALPANAS, 2012).

O método da co-ocorrência, um dos mais comumente empregados na abordagem léxica, é apresentado por Becker e Tumitan (2013), que destacam seu potencial para a análise de *tweets*:

Um dos métodos mais utilizados na abordagem linguística é o da co-ocorrência entre alvo e sentimento, que não leva em consideração nem a ordem dos termos dentro de um documento (*bag-of-words*), nem suas relações léxico-sintáticas. Para a classificação do sentimento em um texto, basta que exista uma palavra de sentimento, cuja polaridade é dada por um léxico de sentimentos. Esse método é extensamente empregado para o atrelamento de um sentimento a uma entidade em uma sentença. [...]. O método por co-ocorrência apresenta bons resultados quando o nível de análise textual é de granularidade pequena, pois a palavra detentora do sentimento está próxima à entidade que qualifica. Sendo assim, este método é usualmente utilizado em análises de nível de sentença, cláusula ou até em documentos com poucos caracteres, como um *tweet* (BECKER e TUMITAN, 2013, p. 11).

Métodos mais avançados podem utilizar uma técnica de processamento de linguagem natural chamada *Part-of-Speech* (POS), em que cada palavra do léxico recebe uma categoria sintática ou morfológica, podendo ser classificada de acordo com sua classe gramatical, como, por exemplo: substantivo, artigo, verbo, adjetivo, advérbio, numeral, pronome, preposição, conjunção ou interjeição (LIU, 2007).

A maior parte dos dicionários léxicos disponíveis é formada exclusivamente por palavras na língua inglesa, como o *Inquirer*, *OpinionFinder*, *SentiWordNet* e *WordNetAffect*, sendo que para língua portuguesa encontram-se o *OpLexicon*

(português do Brasil) e o *SentiLex-PT* (português de Portugal). Há ainda o *Linguistic Inquiry and Word Counts* (LIWC), um software desenvolvido para analisar elementos estruturais, cognitivos e emocionais de documentos textuais, suportando diversos idiomas (BECKER e TUMITAN, 2013).

Tsytsarau e Palpanas (2012) aconselham que a abordagem léxica seja combinada com outras, como a abordagem semântica ou métodos de aprendizados de máquina, uma vez que léxicos de palavras geralmente não são capazes de adaptar valores de polaridade a contextos específicos.

#### 2.3.4.1.2 Abordagem baseada em aprendizado de máquina

Segundo Liu e Zhang (2012), tarefas de classificação de sentimento que utilizam abordagem baseada em aprendizado de máquina geralmente são implementadas por meio de técnicas de aprendizado supervisionado, embora existam algumas técnicas baseadas em métodos não-supervisionados.

Independentemente do método aplicado, o aprendizado de máquina pode ser descrito em linhas gerais por um processo de quatro passos: (1) obtém-se um conjunto de documentos para treinamento, os quais podem ser anotados (aprendizado supervisionado) ou não anotados (aprendizado não supervisionado) com rótulos de sentimento; (2) cada documento é transformado em um vetor de palavras; (3) o classificador é treinado para distinguir fazer uma distinção entre os rótulos de sentimento por meio da análise de cada palavra do vetor; (4) o classificador é utilizado para prever sentimentos de um novo documento, não presente no conjunto de treinamento (TSYRSARAU e PALPANAS, 2012).

O aprendizado de máquina que utiliza métodos supervisionados é explicado por Liu e Zhang por meio de um exemplo bastante ilustrativo (2012):

A classificação de sentimentos obviamente pode ser formulada como um problema de aprendizado supervisionado com três classes: positivo, negativo e neutro. Os dados utilizados em conjuntos de treinamento e de teste usados nas pesquisas são em sua maioria *reviews* de produtos [...]. Uma vez que cada *review* já possui uma nota atribuída pelo usuário (por exemplo, de uma a cinco estrelas), dados para treinamento e teste estão amplamente disponíveis. Por exemplo, um *review* com quatro ou cinco estrelas é considerado positivo, um *review* com uma ou duas estrelas é considerado negativo e um *review* com três estrelas é considerado neutro (LIU e ZHANG, 2012, p. 423, tradução nossa).

Vinodini e Chandrasekaran (2012) apontam que técnicas de aprendizado de máquina como Naive Bayes (NB), *Maximum Entropy* (ME) e *Support Vector Machines* (SVM) tem obtido um grande sucesso na classificação de textos, através dos *métodos K-Nearest Neighbourhood* (KNN), *ID3*, *C5*, *centroid classifier*, *winnow classifier*, e *n-gram model*. De acordo com trabalho realizado por Pang, Lee e Vaithyanathan (2002), a técnica SVM se destacou quanto ao desempenho em relação a ME e NB, o que foi confirmado pelo trabalho de Tsysarau e Palpanas (2012) quando a técnica SVM atingiu uma taxa de acerto superior a 80%, considerando a precisão média da validação cruzada com três “*folds*”, utilizando uma base de teste pré-rotulada contendo *reviews* de filmes, sendo 700 rotulados previamente com sentimento positivo e outros 700 com sentimento negativo.

Apesar disso, Becker e Tumitan (2013) destacam que as melhores taxas de acerto são atingidas quando o conjunto de testes é relativamente esparsa e a rotulagem dos textos é realizada em de forma manual (supervisionada), o que pode ser uma limitação quando o conjunto de dados a ser analisado for amplo. A taxa de acerto é influenciada ainda pela qualidade do conjunto de treinamento e pelo domínio que está sendo analisado.

#### 2.3.4.1.3 Abordagem estatística

Para tentar resolver os problemas referentes à abordagem léxica e baseada em aprendizado de máquina, as abordagens estatísticas buscam identificar a polaridade de uma palavra ao analisar sua co-ocorrência com outras palavras sabidamente positivas ou negativas (TSYT SARAU e PALPANAS, 2012). Isso é realizado com base na premissa de que palavras que traduzem opiniões geralmente são encontradas juntas nos documentos avaliativos; se a palavra ocorre mais frequentemente junto a palavras positivas no mesmo contexto, então ela é provável que seja positiva. De modo análogo, se a palavra possui maior co-ocorrência com palavras sabidamente negativas, então provavelmente seja classificada como negativa. (BECKER e TUMITAN, 2013).

Dessa forma, a polaridade de uma palavra desconhecida pode ser calculada por meio de sua co-ocorrência com qualquer palavra do dicionário que, independentemente de contexto, conserve sua polaridade, como “bom”, por exemplo (TSYT SARAU e PALPANAS, 2012). Para representar isso, Turney (2002) propôs a

adaptação de um conhecido método para o cálculo da dependência estatística, denominado *point-wise mutual information* (PMI), substituindo-se valores de probabilidade pelas frequências do termo de ocorrência, denotado por  $F(x)$  e co-ocorrência, denotados por  $F(x \text{ near } y)$ , representados na seguinte equação (TSYTSARAU e PALPANAS, 2012):

$$PMI(x, y) = \log_2 \frac{F(x \text{ near } y)}{F(x)F(y)} \quad (16)$$

Com base nesse conceito, a polaridade de sentimento, representada por PMI-IR, para uma palavra  $x$  pode ser calculada por meio da diferença entre o valor de PMI computados em relação a um conjunto de palavras com polaridade conhecida positiva ( $pWords$ ) e outro conjunto com polaridade negativa ( $nWords$ ) (TSYTSARAU e PALPANAS, 2012):

$$PMI - IR(x) = \sum_{p \in pWords} PMI(x, p) - \sum_{n \in nWords} PMI(x, n) \quad (17)$$

Becker e Tumitan (2013) aconselham a utilização da abordagem estatística em conjunto com as abordagens léxica e orientadas a aprendizado de máquina, uma vez que são menos sensíveis ao contexto e, dessa maneira, podem contribuir para atingir uma maior taxa de acerto.

#### 2.3.4.1.4 Abordagem semântica

Assim como a abordagem estatística, a determinação da polaridade de documentos de sentimento por meio da semântica é um tipo de análise não supervisionada ao passo que não necessita da rotulação de um conjunto de treinamento para a mineração dos dados (VINHODINI e CHANDRASEKARAN, 2012).

A abordagem semântica sustenta-se no princípio de que palavras semanticamente próximas devem receber valores similares de sentimento; dessa forma, é usualmente utilizada de modo complementar à abordagem léxica, à medida em que, a partir da análise dos sinônimos e antônimos das palavras com polaridade conhecida, é possível inferir a respeito da polaridade das palavras associadas,

permitindo assim uma ampliação dos dicionários existentes (TSYTSARAU e PALPANAS, 2012).

Conforme explicam Becker e Tumitan (2013), essa inferência pode ser feita por meio da mensuração da distância entre palavras usando métricas como caminho ou ainda a partir da contagem da frequência com que as palavras analisadas são associadas a sinônimos positivos ou negativos. O *WordNet*, proposto por Miller et. al. (1990) – um dicionário léxico *online* no qual as palavras são organizadas em conjuntos de sinônimos, que são ligados por meio de diferentes relações, como antonimicidade, por exemplo – é utilizado em vários trabalhos como ponto de partida para a análise semântica (KIM e HOVY, 2004; HU e LIU, 2004; KAMPS et. al., 2004; ESULI e SEBASTIANI, 2005).

#### 2.3.4.2 *Mineração de opinião baseada em características*

Embora a análise da polaridade de textos avaliativos seja útil em muitos casos, pode não ser suficiente em outros. Principalmente em sentenças longas, a classificação de um texto avaliativo como positivo não significa necessariamente que todas as características do objeto de análise agradaram o emissor da opinião; da mesma forma, se o texto for categorizado como negativo, não significa que o avaliador não gostou de nenhuma característica do objeto analisado (BHUIYAN, XU e JOSANG, 2009).

Assim como a classificação de polaridade, a tarefa de mineração de opinião de textos avaliativos funciona também em nível de sentença. Entretanto, para resolver o problema apresentado, ao invés de classificar integralmente a sentença como positiva ou negativa, a tarefa de mineração de opinião baseada em características busca extrair os componentes e os atributos que aparecem na sentença a respeito do objeto analisado, exibindo ao final do processamento um relatório condensado, no qual são apresentados os sentimentos relacionados a cada característica extraída (THORAT, 2014).

Para que isso seja viabilizado, o objeto é representado como uma árvore, cuja raiz é o próprio objeto. Cada nó não-raiz dessa árvore consiste em um componente do objeto que, por sua vez, pode ser associado com um conjunto de atributos. Dessa forma, a opinião pode ser expressa sobre cada nó – raiz ou não – ou sobre cada atributo dos nós. Em uma sentença em que o objeto avaliado é uma câmera digital,

por exemplo, pode conter um conjunto de componentes, como lente, bateria ou tela; a bateria, por sua vez, poderia ter um conjunto de atributos, como duração, tamanho e peso (LIU, 2007).

A partir do exemplo supracitado, poder-se-ia extrair, por meio da mineração de opinião baseada em características, uma avaliação positiva a respeito da duração da bateria e, ao mesmo tempo, uma avaliação negativa a respeito da lente. Essas avaliações parciais seriam combinadas para se chegar a uma avaliação a respeito do objeto como um todo.

#### 2.3.4.3 Mineração relacional

Ao contrário das outras duas tarefas de mineração de opinião, no qual se considera que cada sentença diz respeito a apenas um objeto em particular, a mineração relacional busca identificar sentenças comparativas e, a partir delas, confrontar dois ou mais objetos similares, identificando os objetos e os atributos comparados na sentença (LIU, 2007). O autor apresenta uma definição de sentença comparativa e a forma como a comparação é executada:

Uma sentença comparativa é uma sentença que expressa uma relação baseada nas similaridades ou diferenças entre dois ou mais objetos. A comparação em uma sentença comparativa é geralmente expressa por meio da forma comparativa ou superlativa de um adjetivo ou advérbio. [...] Uma comparação pode ser entre dois ou mais objetos, grupos de objetos, um objeto e demais objetos da mesma categoria. Pode também ser entre um objeto e suas versões anteriores ou futuras (LIU, 2007, p. 433, tradução nossa).

A partir da execução da tarefa de mineração relacional sobre a sentença “A velocidade de processamento do *smartphone* ‘A’ é maior do que do *smartphone* ‘B’” poder-se-ia extrair os objetos comparados (*smartphone* A, *smartphone* B), o componente comparado (processador) e o atributo utilizado na comparação (velocidade) e ainda inferir que esse atributo é superior no aparelho B, de acordo com a opinião comparativa emitida pelo avaliador.

#### 2.3.5 Mineração de opinião no Twitter

A análise de sentimentos no Twitter apresenta algumas características próprias quando comparada a análise realizada em documentos convencionais, como *blogs*,

fóruns de discussão ou *reviews* de produtos. Uma delas consiste na necessidade de contínua adaptação do modelo para analisar os dados, caso esses sejam coletados em tempo real; por outro lado, há uma grande disponibilidade de anotações de sentimento ou humor nas mensagens, o que pode significar uma boa fonte de dados de treinamento para os classificadores (TSYTSARAU e PALPANAS, 2010).

Nos últimos anos, alguns trabalhos vêm sendo desenvolvidos com o objetivo de analisar sentimentos a partir de dados provenientes dessa mídia social. Go, Bhayani e Huang (2009) mostraram que a utilização de anotações de sentimentos para rotulagem dos dados de treinamento é a maneira mais eficaz de realizar aprendizado supervisionado a partir de dados do Twitter. Foram utilizados três diferentes classificadores (NB, ME e SVM), que alcançaram taxas de acerto em torno de 80%.

Pak e Paroubek (2010) utilizaram a API do Twitter para coletar postagens nessa mídia social e empregaram rótulos de sentimento para formar um conjunto de documentos divididos em três classes – sentimentos positivos, sentimentos negativos e textos objetivos (sem sentimentos). Os pesquisadores testaram diferentes classificadores, sendo que o NB (Naive Bayes) apresentou melhores resultados.

Birmingham e Smeaton (2010) também utilizaram a API do Twitter para coletar mais de 60 milhões de *tweets*; o algoritmo de classificação *multinomial* NB (MNB) se sobressaiu ligeiramente em relação ao SVM, alcançando uma taxa de acerto de 74,85%. Os autores também concluíram que a análise de sentimentos em *microblogs*, como o Twitter, obteve melhor taxa de acerto do que a análise de blogs, influenciada pela limitação do número de caracteres e a maior riqueza de mensagens opinativas postadas naquela mídia social.

Tumasjan et. al. (2010) coletaram *tweets* para tentar prever os resultados da eleição federal alemã de 2009. Por meio de uma abordagem léxica, utilizando o software LIWC, 100 mil mensagens que possuem referência a algum partido político ou candidato daquela eleição foram analisadas. Os pesquisadores concluíram que o Twitter é uma plataforma amplamente utilizada para discussões políticas e que há indícios de que a simples contagem de *tweets* reflete a preferência dos usuários, com resultados próximos a pesquisas eleitorais tradicionais.

Bollen, Mao e Zeng (2011) utilizaram dados do Twitter para prever o humor do mercado de ações, com base em métodos estatísticos para análise de sentimentos,

obtendo uma precisão de 87,6% ao comparar os valores obtidos na predição e as cotações diárias de fechamento das ações.

Lima (2012) coletou dados gerados por usuários do Twitter sobre programas de TV. Esses dados foram analisados por meio da aplicação de um classificador supervisionado baseado no NB, que gerou conjuntos de treinamento a partir de três abordagens: rótulos de emoção, palavras e híbrido; foi utilizada também uma ferramenta para desambiguação de sentido baseado em contexto. A precisão obtida na classificação foi de 96,63% para duas classes (positivo, negativo) e 87,25% para três classes (positivo, negativo, neutro).

Miranda Filho (2014) coletou *tweets* para realização de pesquisas de opinião a partir da análise de sentimentos, utilizando abordagens léxicas, semânticas e baseadas em aprendizado de máquina (supervisionado e não supervisionado). Na rotulação manual, obteve-se melhor precisão com a utilização do algoritmo NB (80,5%); na automática, a maior precisão foi obtida com o uso do algoritmo MNB (74,8%). O autor observou ainda que a técnica desenvolvida no trabalho possa ser considerada uma boa alternativa se comparada aos métodos tradicionais de pesquisa.

## 2.4 ALINHAMENTO TEÓRICO

Destacam-se, após apresentação dos elementos teóricos relacionados ao tema estudado, os principais fundamentos que orientaram o desenvolvimento dessa pesquisa e que serviram como base teórica para construção, aplicação e avaliação do modelo proposto:

- a) a equidade da marca (tradução de *brand equity*) pode ser observada sob duas perspectivas: financeira ou do consumidor. Essa pesquisa se interessou em investigar a equidade da marca sob a perspectiva do consumidor, definida, conforme apresenta Keller (1993, p. 8), por meio do “efeito diferencial do conhecimento da marca na resposta do consumidor ao marketing de uma marca” e explicada por Keller e Machado (2006, p. 30): “o *brand equity* está relacionado ao fato de se obterem com uma marca resultados diferentes daqueles que se obteriam se o mesmo produto ou serviço não fosse identificado por aquela marca”;
- b) os elementos formadores da equidade da marca sob a perspectiva do consumidor que foram utilizados para construção do modelo proposto por

essa pesquisa são: lealdade da marca, conhecimento da marca, qualidade percebida e associações à marca;

- c) para avaliar a equidade da marca sob a perspectiva do consumidor a partir das redes sociais virtuais – também conhecidas como redes sociais online ou mídias sociais – as mesmas foram caracterizadas, com base nas definições apresentadas por Elisson e Boyd (2007) e Benevenuto, Almeida e Silva (2011), como um serviço disponibilizado na internet que possibilita aos seus usuários construir perfis públicos ou semipúblicos, manter uma lista de conexões com outros usuários e ainda visualizar e percorrer essa lista de conexões;
- d) anteriormente ao surgimento das redes sociais virtuais, sociólogos já se interessavam no estudo das relações entre entidades sociais, desenvolvendo, ao longo do tempo, uma série de técnicas para a investigação dessas redes, dando origem a uma área de investigação denominada análise de redes sociais (ARS). A ARS fez o uso de várias fundamentações da teoria dos grafos para propor um conjunto de métricas que caracterizassem as redes sociais e seus elementos, tornando possível a comparação entre diferentes redes sociais, a identificação de mudanças na rede em função do tempo e ainda a determinação da importância de indivíduos ou grupos dentro de uma rede. No modelo proposto por essa pesquisa, utilizou-se uma métrica denominada *PageRank*, apresentada pelos criadores do Google Sergey Brian e Larry Page. Inicialmente proposta com o objetivo de identificar o prestígio de páginas web em uma rede, seu conceito foi expandido para determinação do prestígio de cada elemento em qualquer rede social, a partir das características de suas ligações com outros elementos da rede;
- e) o processo de descoberta de conhecimento, tradução de *Knowledge Discovery in Databases* é um processo não trivial para a extração de informação implícita, previamente desconhecida e potencialmente útil, a partir dos dados (FRAWLEY, PIATETSKY-SHAPIRO e MATHEUS, 1992). A mineração de dados é uma etapa importante desse processo, que consiste na aplicação de algoritmos específicos para a extração de padrões a partir desses dados, geralmente empregando aprendizagem de máquina, reconhecimento de padrões e técnicas estatísticas (FAYYAD, PIATETSKY-

SHAPIRO e SMYTH, 1996). As tarefas de mineração de dados podem ser divididas em preditivas (classificação e regressão) ou descritivas (sumarização, associação, agrupamento e detecção de anomalias). Com objetivos diferentes, dois algoritmos relacionados a técnicas de classificação foram empregados no modelo proposto por essa pesquisa: o SVM (*Support Vector Machine*), em sua implementação JMySVM, empregado nessa pesquisa para a classificação de sentimentos e o algoritmo C4.5, em sua implementação J48, para a determinação do tipo de cada perfil coletado da rede social Twitter;

f) a análise de sentimentos consiste no “estudo computacional de opiniões, sentimentos e emoções expressas em texto” (MOENS, 2009). A tarefa de mineração de sentimentos empregada na pesquisa foi a classificação de sentimentos, também conhecida como classificação de polaridade, cujo objetivo é classificar um texto avaliativo como positivo ou negativo. Nessa pesquisa, a classificação de sentimentos foi empregada para atribuir polaridade positiva ou negativa aos textos publicados por usuários da rede social Twitter.

Assim sendo, essa pesquisa realizou o emprego dos fundamentos supracitados para avaliar a equidade da marca a partir de dados coletados das redes sociais virtuais. A aplicação desses embasamentos teóricos segue a metodologia apresentada na Seção 3: Metodologia da Pesquisa.

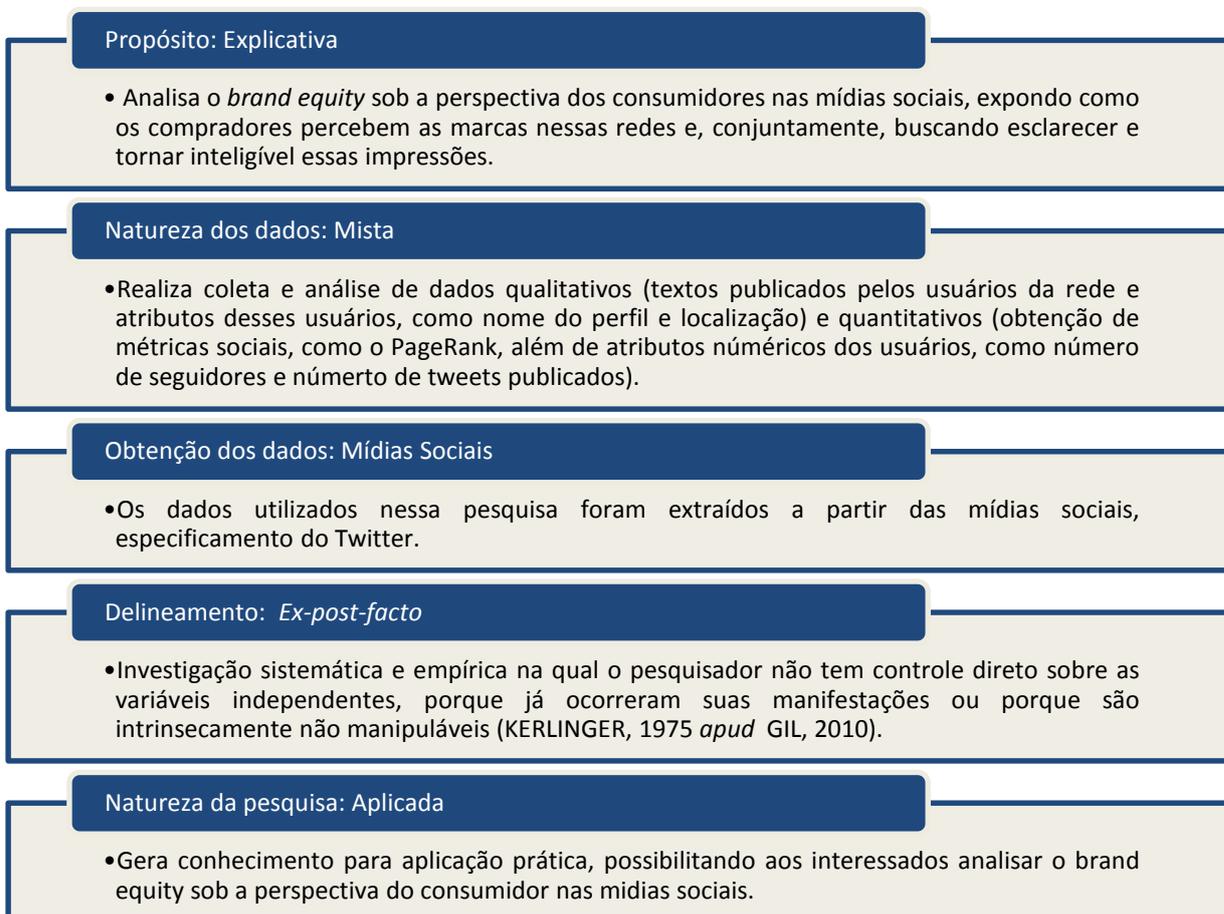
### 3 METODOLOGIA DA PESQUISA

O presente capítulo estabelece a metodologia de pesquisa adotada para esse trabalho, por meio da tipificação da pesquisa com base em suas características, da descrição do ambiente de pesquisa e dos procedimentos metodológicos adotados para que os objetivos específicos sejam atingidos.

#### 3.1 CARACTERIZAÇÃO DA PESQUISA

Vergara (1997) propôs uma classificação de pesquisa com base em cinco critérios: quanto ao propósito, quanto à natureza dos dados, quanto ao delineamento, quanto à forma de obtenção dos dados e quanto à natureza da pesquisa. A Figura 23 sintetiza a caracterização da presente pesquisa de acordo com os critérios propostos pela autora.

**FIGURA 23 – CARACTERIZAÇÃO DA PESQUISA QUANTO AO PROPÓSITO, NATUREZA DOS DADOS, FORMA DE OBTENÇÃO DOS DADOS, DELINEAMENTO E NATUREZA DA PESQUISA**



FONTE: O AUTOR (2015)

A pesquisa desenvolvida nesse trabalho pode ser classificada quanto ao propósito como explicativa, na medida em que busca analisar o *brand equity* sob a perspectiva dos consumidores nas mídias sociais, expondo como os compradores percebem as marcas nessas redes e, conjuntamente, esclarecendo e tornando inteligíveis. Quanto à natureza dos dados, trata-se de uma pesquisa mista, tendo em conta que coleta e analisa tanto dados qualitativos quanto quantitativos, obtidos a partir das mídias sociais, especificamente do Twitter. Em relação ao delineamento da pesquisa, trata-se de uma pesquisa *ex-post-facto*; quanto à natureza, trata-se de uma pesquisa aplicada, uma vez que busca gerar conhecimentos para aplicação prática, possibilitando aos interessados, sobretudo profissionais da área de marketing, a descoberta de conhecimento a partir das redes sociais.

### 3.2 AMBIENTE DA PESQUISA

A pesquisa desenvolvida neste trabalho tem as redes sociais como âmbito, especificamente no Twitter. Os usuários desse serviço de microblog – constituído por pessoas comuns das mais distintas classes sociais, celebridades, políticos e até mesmo chefes de estado de diversos países – escrevem mensagens de texto curtas sobre fatos cotidianos e compartilham opiniões sobre uma grande gama de tópicos, expressando seus sentimentos em relação a visões políticas ou religiosas ou ainda sobre produtos e serviços que eles utilizam. Essas peculiaridades permitem que os dados recuperados a partir dessa mídia social possam ser eficientemente utilizados para o marketing ou estudos sociais, servindo de insumo para tarefas de análise de sentimentos (PAK e PAROUBEK, 2010).

Os dados foram coletados por meio da ferramenta NodeXL (descrita na seção 3.3.1), que é executada sobre a API disponibilizada pelo próprio Twitter. API é um acrônimo para *Application Programming Interface* ou Interface de Programação de Aplicativos, em português. (CIRIACO, 2009). A API do Twitter provê acesso à funções pré-programadas que possibilitam a leitura e escrita de dados na rede social; essas funções, quando acessadas por intermédio de rotinas computacionais, permitem, por exemplo, publicar *tweets*, ler o perfil dos autores dos *tweets* ou buscar os seguidores de determinado perfil (TWITTER, 2014).

### 3.3 FERRAMENTAS

Nas subseções 3.3.1, 3.3.2, 3.3.3 e 3.3.4 são descritas as ferramentas que deram suporte ao desenvolvimento dessa pesquisa.

#### 3.3.1 NodeXL

O NodeXL (<http://nodexl.codeplex.com>) é uma ferramenta livre e de código aberto integrada ao Microsoft Excel (2007, 2010 e 2013) e mantida pela *Social Media Research Foundation* (NODEXL, 2015). Essa ferramenta provê um conjunto de funcionalidades que auxiliam os processos de análise e visualização de redes sociais, incluindo ainda suporte para coleta de dados a partir de mídias sociais, como *Facebook*, *Flickr*, *Twitter* e *Youtube*. Para isso, utiliza um modelo estruturado de planilha que contém abas para o armazenamento de informações necessárias para representação de uma rede (HANSEN, SHNEIDERMAN e SMITH, 2010).

Assim, os relacionamentos de uma rede – ou as arestas de um grafo – são representadas em uma aba denominada “*Edges*”, que contém todos os pares de vértices que estão conectados na rede, bem como alguns atributos complementares (HANSEN, SHNEIDERMAN e SMITH, 2010; NODEXL, 2015). Quando essas arestas são construídas a partir da coleta de mensagens da rede social Twitter, são disponibilizados pela ferramenta 14 atributos ao todo, detalhados no Quadro 2.

**QUADRO 2 – DESCRIÇÃO DOS ATRIBUTOS RELACIONADOS AOS TWEETS DISPONIBILIZADOS PELA FERRAMENTA NODEXL**

<b>Atributo</b>	<b>Descrição</b>
<i>Relationship</i>	<p>A ferramenta NodeXL classifica os <i>tweets</i> buscados em três tipos de relacionamento: <i>Tweet</i>, <i>Replies To</i> e <i>Mentions</i>.</p> <p>Se um <i>tweet</i> é uma resposta direcionada a algum usuário, o NodeXL cria uma aresta partindo do usuário que publicou a resposta para o usuário que foi respondido, atribuindo a esse relacionamento o valor "<i>Replies To</i>". Essas mensagens também são conhecidas como dirigidas, apresentadas na seção 2.2.6 dessa pesquisa.</p> <p>Se o <i>tweet</i> menciona algum outro usuário, o NodeXL cria uma aresta partindo do usuário que mencionou para o usuário mencionado, atribuindo a esse relacionamento o valor "<i>Mentions</i>". Essas mensagens também são conhecidas como <i>retweet</i>, apresentadas na seção 2.2.6 dessa pesquisa.</p> <p>Se o <i>tweet</i> não foi nem uma resposta, nem uma menção, o NodeXL cria um laço do tipo loop partindo do usuário que publicou e apontando para ele próprio, atribuindo à relação o valor "<i>Tweet</i>". Essas mensagens podem ser chamadas de não direcionadas, apresentadas na seção 2.2.6 dessa pesquisa.</p>
<i>Vertex1</i>	Nome do perfil do usuário que publicou o <i>tweet</i>
<i>Vertex2</i>	Nome do perfil do usuário secundário relacionado ao <i>tweet</i> , podendo ser o usuário respondido (no caso da relação <i>Replies To</i> ), o usuário retuítdo (no caso da relação <i>Mentions</i> ) ou o próprio autor do <i>tweet</i> (no caso da relação do tipo <i>Tweet</i> ).
<i>Relationship Date (UTC)</i>	Data/hora em que o relacionamento foi constituído, no fuso horário UTC.
<i>Tweet</i>	Texto do <i>tweet</i> coletado
<i>URLs in Tweet</i>	Urls presentes no <i>tweet</i> coletado
<i>Domains in Tweet</i>	Domínios web presentes do <i>tweet</i> coletado
<i>Hashtags in Tweet</i>	Hashtags presentes no <i>tweet</i> coletado (significado explicado na seção 2.2.6 dessa pesquisa).
<i>Tweet Date (UTC)</i>	Data/hora de publicação do <i>tweet</i> , no fuso horário UTC. Coincide com o atributo <i>Relationship Date (UTC)</i> .
<i>Twitter Page for Tweet</i>	Link do Twiter que direciona ao <i>tweet</i> coletado
<i>Latitude</i>	Latitude em que o <i>tweet</i> coletado foi publicado, quando disponível.
<i>Longitude</i>	Longitude em que o <i>tweet</i> coletado foi publicado, quando disponível.
<i>Imported ID</i>	Identificador único do <i>tweet</i> coletado

<i>In-Reply-To Tweet ID</i>	Identificador único do tweet respondido, caso do relacionamento seja do tipo " <i>Replies To</i> ".
-----------------------------	---

FONTE: O AUTOR (2015)

Além dessa, existem outras abas na planilha que possuem informação a respeito de cada vértice (aba "Vertices") – como, por exemplo, rótulo do vértice, propriedades visuais e métricas relacionadas (HANSEN, SHNEIDERMAN e SMITH, 2010; NODEXL, 2015). Quando as mensagens são coletadas pela ferramenta a partir da rede social Twitter, os dados dos usuários que publicaram as mensagens são armazenados nessa aba. O Quadro 3 descreve os atributos disponibilizados pela ferramenta relacionados a esses usuários.

**QUADRO 3 – DESCRIÇÃO DOS ATRIBUTOS DISPONIBILIZADOS PELA FERRAMENTA NODEXL REFERENTES AOS USUÁRIOS DA REDE SOCIAL TWITTER**

<b>Atributo</b>	<b>Descrição</b>
<i>Vertex</i>	Nome de perfil do usuário que publicou o <i>tweet</i> , do usuário respondido pelo <i>tweet</i> ou do usuário retuítdo.
<i>Image File</i>	Link para a imagem do perfil do usuário
<i>Followed</i>	Número de perfis seguidos pelo usuário
<i>Followers</i>	Número de perfil que seguem o usuário
<i>Tweets</i>	Número de <i>tweets</i> publicados pelo usuário
<i>Favorites</i>	Número de <i>tweets</i> marcados como "favorito" pelo usuário
<i>Time Zone UTC Offset (Seconds)</i>	Diferença, em segundos, do fuso horário do usuário para o fuso horário de referência (UTC), quando disponível.
<i>Description</i>	Descrição do perfil fornecida pelo usuário, quando disponível.
<i>Location</i>	Localização fornecida pelo usuário, quando disponível.
<i>Web</i>	Página da web fornecida pelo usuário, quando disponível.
<i>Time Zone</i>	Fuso horário do usuário, quando disponível.
<i>Joined Twitter Date (UTC)</i>	Data/hora de criação do perfil, no fuso horário UTC.
<i>Custom Menu Item Text</i>	Local de onde o <i>tweet</i> foi coletado. Como todos os <i>tweets</i> coletados para essa pesquisa são públicos, o valor desse campo sempre será " <i>Open Twitter Page for This Person</i> ".
<i>Tweeted Search Term?</i>	Apresenta o valor "Yes" caso o usuário tenha publicado o termo buscado e "No", caso contrário.

FONTE: O AUTOR (2015)

Existem ainda as seções "*Group*" e "*Group Vertices*", referentes aos agrupamentos da rede e ainda outra aba denominada "*Overall Metrics*", que apresenta

algumas métricas relacionadas à rede como um todo, como o tipo da rede (direcionada ou não direcionada), número de vértices, número de arestas, número de componentes conectados, número de componentes que contém apenas um vértice, número máximo de vértices em um componente conectado, número máximo de arestas em um componente conectado, diâmetro da rede (distância geodésica máxima), distância geodésica média e densidade da rede (HANSEN, SHNEIDERMAN e SMITH, 2010; NODEXL, 2015).

### 3.3.2 Rapidminer

*Rapidminer* (<https://rapidminer.com>) é uma plataforma livre, adaptável e de código aberto desenvolvida na linguagem Java que provê um ambiente integrado para análise de dados, por meio de soluções baseadas em aprendizado de máquina, mineração de dados, mineração de textos e técnicas estatísticas (BURGET et. al., 2010; HOFMANN e KLINKENBERG, 2013).

Essa ferramenta emprega uma abordagem baseada na utilização de operadores modulares, cada qual com sua função específica, que podem ser concatenados em cadeias complexas, as quais são arranjadas de acordo com o problema a ser resolvido; a saída de cada operador é utilizada como entrada para o próximo, e assim sucessivamente, até que o processamento seja finalizado (GRACZYK, LASOTA e TRAWIŃSKI, 2009).

### 3.3.3 Weka

Weka (<http://www.cs.waikato.ac.nz/ml/weka>), um acrônimo para *Waikato Environment for Knowledge Analysis*, é um projeto não-comercial e de código aberto que fornece suporte para algumas etapas relacionadas ao processo de descoberta de conhecimento, como pré-processamento, mineração de dados (por meio das tarefas de classificação, associação e agrupamento) e visualização (HALL et. al. 2009; GRACZYK, LASOTA e TRAWIŃSKI, 2009; RAMAMOCHAN et. al. 2012).

O projeto teve início em 1992, quando algoritmos de aprendizado de máquina estavam disponíveis em diferentes plataformas e os formatos de entrada e saída de dados não eram padronizados. O WEKA foi então idealizado como uma solução para esse problema, ao prover não apenas uma caixa de ferramentas esses algoritmos,

mas um ambiente por meio do qual pesquisadores poderiam implementar novos códigos sem se preocupar com a infraestrutura para a preparação e manipulação dos dados. Originalmente escrito na linguagem C, o WEKA foi posteriormente completamente reescrito na linguagem Java, tornando-se compatível com a maioria das plataformas computacionais. Atualmente, o WEKA é uma referência como sistema para mineração de dados e aprendizagem de máquina, alcançando ampla aceitação nos meios acadêmicos e empresariais (HALL et. al. 2009, RAMAMOCHAN et. al. 2012).

#### 3.3.4 Tagul

O Tagul (<https://tagul.com>) é uma ferramenta online utilizada para análise e visualização de dados textuais, por meio da apresentação de dados estatísticos relacionados à frequência com que as palavras ou frases aparecem em um texto e da geração de nuvens de palavras, uma representação gráfica em que há uma proporção direta entre a frequência das palavras no texto e o tamanho da palavra no conglomerado gerado (SANTOS, DE PRÓSPERO e GIANORDOLI, 2014).

#### 3.3.5 DD-CSS

Abreviação para *Data-driven Computational Social Science*, DD-CSS (<http://dd-css.com>) é uma aplicação *web*, de código aberto, desenvolvida em Python, que tem como intuito auxiliar pesquisadores a coletar e analisar dados a partir de mídias sociais. Possibilita, por exemplo, obter identificadores de seguidores e amigos de um determinado perfil do Twitter ou ainda coletar até 3200 *tweets* de qualquer perfil específico; permite ainda a consulta ao número de compartilhamento de uma determinada url no Facebook. Os arquivos de saída podem ser obtidos nos formatos JSON ou CSV (DD-CSS, 2014).

### 3.4 PROCEDIMENTOS METODOLÓGICOS

No Quadro 4 são apresentados os procedimentos metodológicos adotados de modo a atender cada objetivo específico dessa pesquisa, relacionando ainda os conceitos e principais autores.

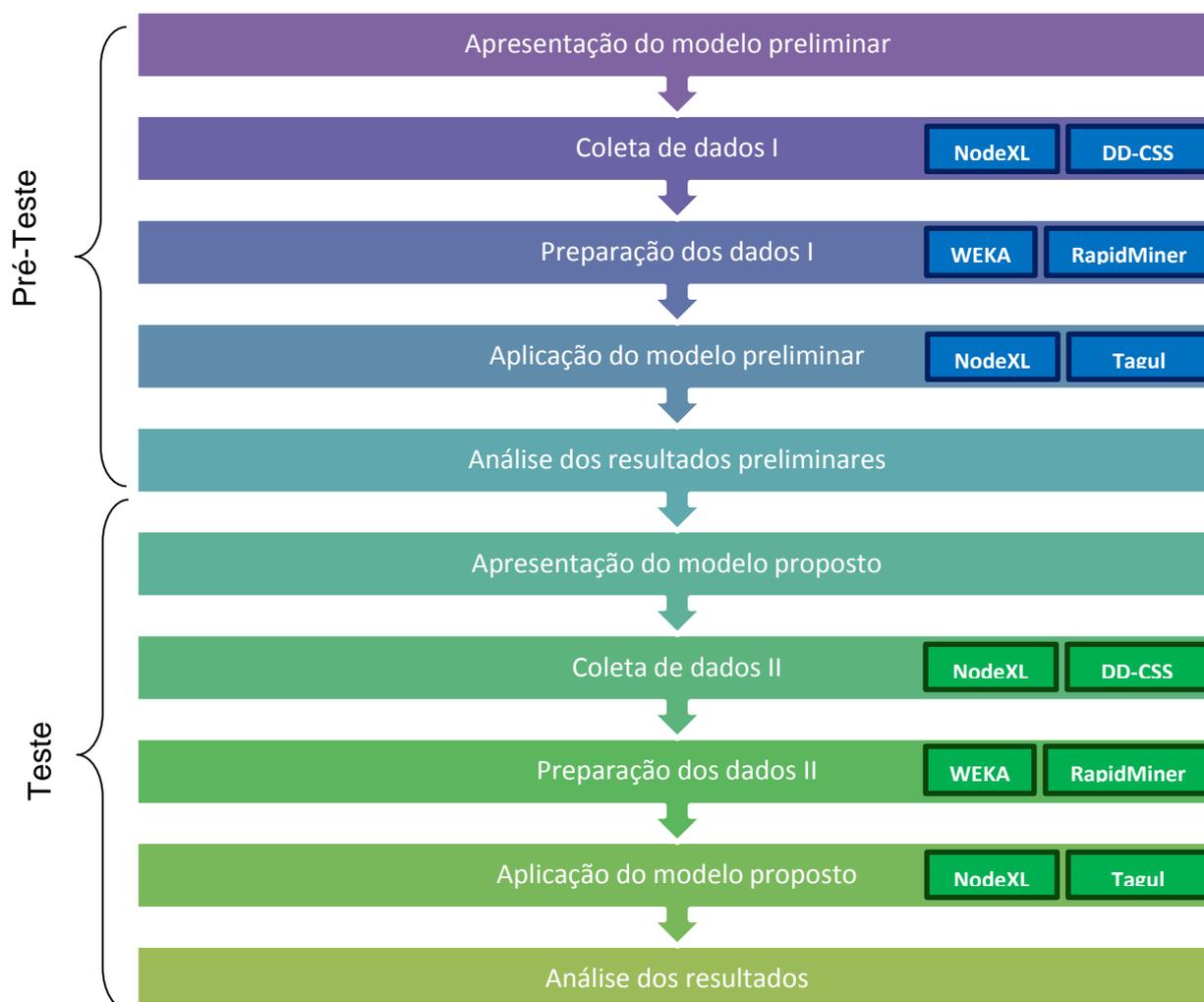
**QUADRO 4 – PROCEDIMENTOS METODOLÓGICOS ADOTADOS DE ACORDO COM OS OBJETIVOS ESPECÍFICOS**

Objetivos específicos	Conceitos	Principais autores	Descrição do Procedimento
Identificar as dimensões presentes nos modelos conceituais de <i>brand equity</i> sob a perspectiva do consumidor;	Marcas, Equidade de Marca e Modelos Conceituais de Equidade de Marca	Keller (1993) Aaker (1998) Keller e Machado (2006) Kotler e Keller (2006) Kapferer (2012) Christodoulides e De Chernatony (2010) Yoo e Donthu (2001)	Levantar os modelos conceituais por meio de pesquisa bibliográfica, utilizando bases internacionais e nacionais de trabalhos acadêmicos; identificar e compreender os critérios de avaliação de <i>brand equity</i> a partir dos modelos encontrados.
Propor um modelo para análise do <i>brand equity</i> sob a perspectiva do consumidor a partir das mídias sociais, contemplando as dimensões identificadas nos modelos conceituais;	Análise de Redes Sociais, Redes Sociais Virtuais	Wasserman e Faust (1994), Benevenuto, Almeida e Silva (2011) Elisson (2007)	Apresentar um modelo com os critérios utilizados para avaliação do <i>brand equity</i> com base nos dados provenientes das redes sociais.
Validar o modelo proposto por meio de aplicação sobre dados gerados por usuários da rede social Twitter;	Descoberta de Conhecimento, Mineração de Dados, Mineração de Opinião	Piatetsky-Shapiro (1990), Frawley, Piatetsky-Shapiro e Mathues (1992), Fayyad, Piatetsky-Shapiro e Smyth (1996) Liu (2007, 2010, 2012) Pang e Lee (2008) Boiy e Moens (2009) Vinodhini e Chandrasekaran(2012)	Coleta: Extração de dados da rede social Twitter por meio da API de busca disponibilizada por essa mídia social.  Análise: Análise de Redes Sociais; Mineração de Opinião.

FONTE: O AUTOR (2014)

Assim, para que os objetivos específicos (a), (b) e (c) e, conseqüentemente, o objetivo geral, fossem atingidos, foram empregados os procedimentos descritos no Quadro 4, aplicados por meio dos passos e ferramentas apresentados na Figura 24.

**FIGURA 24 – PASSOS EXECUTADOS E FERRAMENTAS UTILIZADAS DURANTE AS ETAPAS DE TESTE E PRÉ-TESTE PARA ALCANCE DOS OBJETIVOS PROPOSTOS PARA A PESQUISA**



**FONTE: O AUTOR (2015)**

Como se pode perceber pela Figura 24, a pesquisa é dividida em duas etapas principais: pré-teste e teste. O pré-teste consiste em uma etapa piloto que busca fornecer subsídios para o aprimoramento do modelo preliminar para análise da equidade de marca a partir dos dados coletados da rede social Twitter. Na etapa de pré-teste, foram analisados *tweets* relacionados a quatro conhecidas marcas de

cosméticos vendidos no Brasil. Assim, com base na análise dos resultados preliminares obtidos, uma evolução do modelo é apresentada, sendo aplicada novamente em dados coletados do Twitter, dessa vez sobre *tweets* relacionados às marcas de montadoras de automóveis.

### 3.4.1 Pré-teste

A etapa de pré-teste consiste na proposta do modelo preliminar, na coleta de dados, preparação dos dados, aplicação do modelo preliminar e análise dos resultados preliminares. As etapas de coleta e preparação de dados são descritas nas subseções 3.4.1.1 e 3.4.1.2, enquanto que o modelo preliminar e a análise dos resultados preliminares são apresentados nas seções 4.1 e 4.2.

#### 3.4.1.1 Coleta de Dados I

Entre 04/03/2014 e 04/05/2015 foram realizadas nove coletas de *tweets* em português sobre quatro marcas de cosméticos presentes no Brasil que possuem perfil oficial em português: Avon ([twitter.com/avonbr](https://twitter.com/avonbr)), O Boticário ([twitter.com/oboticario](https://twitter.com/oboticario)), Jequiti ([twitter.com/jequiti](https://twitter.com/jequiti)) e Natura ([twitter.com/naturanet](https://twitter.com/naturanet)). O critério utilizado para a escolha dessas marcas foi não-probabilístico e intencional. Essas quatro marcas foram utilizadas no pré-teste dessa pesquisa.

As coletas foram realizadas por meio da ferramenta NodeXL. Devido às limitações da API de busca do Twitter, que é utilizada pela ferramenta para a recuperação dos dados, a janela permitida para pesquisa é de sete dias, aproximadamente. Assim sendo, a primeira coleta, realizada no dia 04/03/2014, buscou *tweets* publicados entre 25/02/2015 e a data da coleta. Deste modo, ao longo dos dois meses de coleta, foram recuperados 177.685 *tweets* de 124.564 usuários, publicados entre 25/02/2015 e 04/05/2015. A consulta executada, bem como a quantidade de *tweets* e usuários retornados são discriminados na Tabela 1.

**TABELA 1 – CONSULTAS REALIZADAS ENTRE 04/03/2015 E 04/05/2015 E NÚMERO DE TWEETS E USUÁRIOS RETORNADOS PARA CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS PESQUISADAS**

Marcas	Consulta	Tweets	Usuários
Avon	"avon lang:pt"	29.353	27.618
Boticário	"boticario lang:pt"	12.834	10.931
Jequiti	"jequiti lang:pt"	71.041	55.759
Natura	"natura lang:pt"	64.457	30.256
Total	N/A	177.685	124.564

FONTE: O AUTOR (2015)

Nessa consulta, foram recuperados ainda 28 atributos complementares referentes aos Tweets e Usuários, disponibilizados pela ferramenta NodeXL e descritos previamente na seção 3.3.1.

Para formar o conjunto de treinamento a ser empregado na classificação de sentimentos dos *tweets* referentes as quatro marcas de cosméticos utilizadas no pré-teste, foram coletados, em 08/05/2015, em uma única consulta, 22.255 tweets em português, com sentimento positivo e negativo, discriminados na Tabela 2.

**TABELA 2 – NÚMERO DE TWEETS COLETADOS PARA FORMAÇÃO DO CONJUNTO DE TREINAMENTO DA ANÁLISE DE SENTIMENTOS (AGRUPADOS POR POLARIDADE ATRIBUÍDA)**

Sentimento	Consulta	Tweets
Positivo	":) lang:pt"	12.421
Negativo	":( lang:pt"	9.834
Total	N/A	22.255

FONTE: O AUTOR (2015)

Por fim, foi necessário coletar ainda os seguidores dos perfis oficiais de cada uma das marcas utilizados no pré-teste, uma vez que isso foi um dos insumos utilizados para a análise do *brand equity* dessas marcas. Mais uma vez por limitações da API de busca do Twitter, foi possível recuperar apenas os identificadores numéricos desses seguidores, por meio da ferramenta DD-CSS. A coleta foi realizada em 11/05/2015 e, portanto, representa os seguidores dos perfis naquele momento. A Tabela 3 mostra um resumo do quantitativo de identificadores coletados de cada um dos perfis analisados.

**TABELA 3 – NÚMERO DE SEGUIDORES OBTIDOS A PARTIR DA COLETA DE CADA UM DOS PERFIS ASSOCIADOS ÀS MARCAS DE COSMÉTICOS ANALISADAS**

Marcas	Perfil	Seguidores
Avon	twitter.com/avonbr	96.387
Boticário	twitter.com/oboticario	98.657
Jequiti	twitter.com/jequiti	32.513
Natura	twitter.com/naturanet	65.183
Total	N/A	292.740

**FONTE: O AUTOR (2015)**

Após a coleta dos *tweets* para formação da base de estudo e da base de treinamento e ainda dos identificadores de cada um dos seguidores dos perfis oficiais das marcas em estudo, foi realizada a preparação dos dados para que pudessem, posteriormente, ser utilizados como entrada para o modelo proposto de avaliação do *brand equity* a partir dos dados coletados das redes sociais. A preparação dos dados é descrita na subseção 3.5.

#### 3.4.1.2 *Preparação dos Dados I*

A fim de que os dados coletados pudessem ser utilizados como artefato de entrada para o modelo proposto nessa pesquisa, os mesmos precisavam ser previamente preparados. Essa preparação foi realizada seguindo-se os cinco passos descritos na Figura 25.

**FIGURA 25 – PASSOS EXECUTADOS NO PRÉ-TESTE DURANTE A ETAPA DE PREPARAÇÃO DOS DADOS**

**FONTE: O AUTOR (2015)**

Cada um dos passos será detalhado nas subseções a seguir.

#### *3.4.1.2.1 União dos conjuntos de tweets coletados*

Cada consulta realizada por meio da ferramenta NodeXL retornou um arquivo com os *tweets* recuperados, totalizando 36 arquivos após a execução das nove buscas para cada marca. Esses arquivos foram manualmente concatenados, resultando em quatro arquivos, cada um referente a uma das quatro marcas utilizadas no pré-teste.

#### *3.4.1.2.2 Deduplicação dos tweets*

Uma vez que a janela de busca da API do Twitter é aproximada, a execução do procedimento descrito no passo anterior poderia resultar em *tweets* duplicados, na medida em que os últimos *tweets* retornados na coleta da semana 1 poderiam coincidir com os primeiros *tweets* retornados na coleta da semana 2, e assim sucessivamente.

A Figura 26 ilustra esse processo. Nela, são apresentados recortes de duas coletas referentes à marca “Avon”, realizadas em 06/04/2015 e 13/04/2015. A primeira delas recuperou *tweets* publicados na janela entre 30/03/2015 às 15h17min e 07/04/2015 às 03h05min, enquanto que a segunda compreende *tweets* publicados

entre 05/04/2015 às 22h35min e 13/04/2015 às 17h49min. Há, portanto, uma interseção de datas resultante das duas consultas, especificamente entre as 22h35min do dia 05/04/2015 (início da janela da segunda consulta) e às 03h05min do dia 07/04/2015 (fim da janela da primeira consulta).

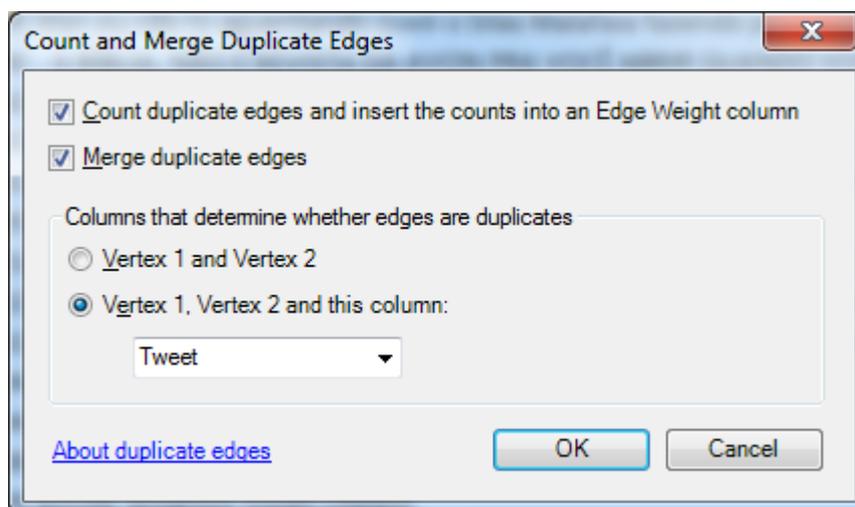
**FIGURA 26 – EXEMPLOS DE TWEETS DUPLICADOS RESULTANTES DE COLETAS SUBJACENTES**

2	Vertex 1	Vertex 2	Relationship	Tweet
3	frnkoceanz	onlyforlesbians	30/03/2015 15:17	RT @onlyforlesbians: "a biblia n é revista da avon p abrir só p pedir" Avon ao menos entrega as coisas, j
4	marcoscezar07	marcoscezar07	30/03/2015 15:22	Minha mãe quer que eu fique acordado esperando avon 🙄🙄
5	beeshopi	beeshopi	30/03/2015 15:32	eu não nasci pra esse mundo eu vou largar tudo e vender avon
⋮				
2206	justeen_life	justeen_life	05/04/2015 22:35	O perfume que eu quero compra na Avon tá 119 .. mais ele é muito cheiroso ..
2207	menestrino	lucianagenro	05/04/2015 22:42	Tuitaço da @lucianagenro contra a redução da maioridade penal... Caras...vão ler catálogo da #Avon q
2208	lucianammts	lucianammts	05/04/2015 22:42	Sobre a prova de amanhã? Nem abri o livro pra estudar ainda 🙄🙄 não decidi ainda se vou vender avo
⋮				
2781	lorenaknust	allanlac	07/04/2015 03:04	RT @allanlac: te dei tudo dei perfume dei rolon te dei tudo dei natura dei avon e vc nada NADA DE FIC
2782	oigilovely	oigilovely	07/04/2015 03:05	sério netflix pra q botar q a mulher q adotou o edward é vendedora da avon
2783	perfumaria_avon	perfumaria_avon	07/04/2015 03:05	curti <a href="http://t.co/4a6kzIU8q">http://t.co/4a6kzIU8q</a>
2784				
⋮				
2	Vertex 1	Vertex 2	Relationship	Tweet
3	justeen_life	justeen_life	05/04/2015 22:35	O perfume que eu quero compra na Avon tá 119 .. mais ele é muito cheiroso ..
4	menestrino	lucianagenro	05/04/2015 22:42	Tuitaço da @lucianagenro contra a redução da maioridade penal... Caras...vão ler catálogo da #Avon que
5	lucianammts	lucianammts	05/04/2015 22:42	Sobre a prova de amanhã? Nem abri o livro pra estudar ainda 🙄🙄 não decidi ainda se vou vender avon
⋮				
2801	_prudencio99	trouxavcs	13/04/2015 17:44	RT @trouxavcs: Milena: vai ser vendedora avon
2802	vinicinbeta	trouxavcs	13/04/2015 17:45	RT @trouxavcs: Milena: vai ser vendedora avon
2803	catiamcosta	catiamcosta	13/04/2015 17:49	Já disse q vem aí review? É.#makeup #lipstick #avon #reviews #amarinar #GIRLYSTUFF #beauty #cosmeti
2804				

**FONTE: O AUTOR (2015)**

Assim, quando a união dessas duas coletas é realizada, todos os tweets publicados nessa janela de interseção são duplicados. Para contornar essa situação, foi utilizada a função “*Count and Merge Duplicate Edges*”, disponibilizada pelo próprio NodeXL. Como parâmetros, foram passadas as seguintes colunas: Vertex 1, Vertex 2 e Tweet, conforme Figura 27.

FIGURA 27 – EXECUÇÃO DA FUNÇÃO *COUNT AND MERGE DUPLICATE EDGES*, EXECUTADA PARA REMOÇÃO DOS TWEETS DUPLICADOS



FONTE: O AUTOR (2015)

Dessa forma, todos os registros duplicados seriam combinados em um único registro, caso o valor desses três atributos fosse o mesmo para diferentes entradas. Além disso, por meio desse procedimento, uma coluna “*Edge Weight*” é inserida na planilha e indica o número de vezes que o *tweet* aparecia antes da deduplicação.

Um efeito secundário, mas não menos importante, resultante da execução da combinação de registros duplicados é a remoção de spams. Spams são mensagens idênticas postadas repetidamente por um mesmo perfil. Com a deduplicação, mesmo que o usuário realize a postagem de uma mesmo *tweet* diversas vezes, apenas uma será considerada para a análise.

A Tabela 4 mostra um comparativo entre a quantidade de *tweets* antes e após a deduplicação.

TABELA 4 – COMPARATIVO ENTRE A QUANTIDADE DE TWEETS ANTES E APÓS A DEDUPLICAÇÃO, REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS

Marcas	Tweets	
	Antes	Após
Avon	29.353	25.434
Boticário	12.834	8.062
Jequiti	71.041	58.603
Natura	64.457	53.222
Total	177.685	145.321

FONTE: O AUTOR (2015)

### 3.4.1.2.3 Deduplicação de usuários

Assim como os *tweets*, usuários duplicados também podem resultar do processo de união dos arquivos obtidos na coleta, como ilustra a Figura 28.

FIGURA 28 – EXEMPLOS DE USUÁRIOS DUPLICADOS RESULTANTES DE COLETAS SUBJACENTES

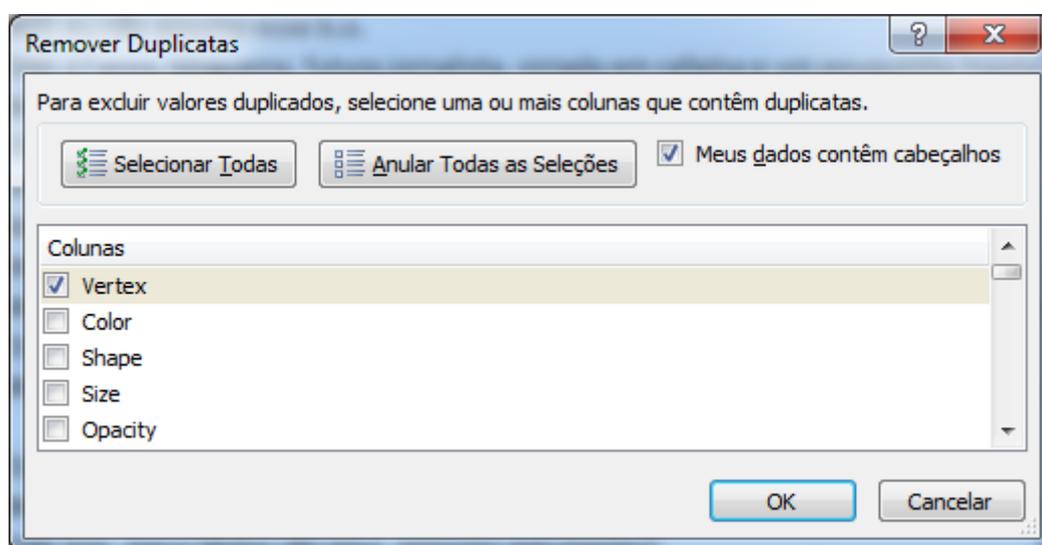
2	Vertex	Followed	Followers	Tweets	Favorites	Time Zone UTC Offset (Seconds)	Description
3	_aline_silva_		479	300	6479	991	é torto, completamente confuso, mas é o meu jeito
4	_carvalhais		502	916	41930	7646	Procure a sua PaZ !!!
5	_johnlennon		187	501	27326	248	-10800 @TheBeatles ♥
6	_nana		294	1457	84133	11018	Ich möchte ein Fennek sein. #MiaSanMia
7	_putspaula		48	364	27754	9128	-7200 Snap: paula_0007 // Insta: oops_paula
8	_70mipegarr		791	1100	5759	884	Corinthians e Real Madrid ☑
9	_agathapaiva		30	185	211	43	sonhar, viver e todo o dia agradecer.
10	_analemos0		179	112	1965	652	A cura para todos os males <span style="float: right;">Está em nós mesm</span>
11	_andradejess		268	674	35528	5230	-10800 você foi minha vida, e eu apenas um capítulo da sua
12	_annaferreira		211	708	57901	1297	-10800
13	_annapqsim		481	821	9363	2711	-14400 maybe you're the rainbow
14	_atst		236	445	26739	8104	-10800 Nunca tem pizza no meu tomate
15	_aumenda		286	183	2641	366	-10800 Leonina ⚡, addicted TV series(vampire manic and Stelena), animals and futur
16	_bekobashigawa		278	318	6147	2929	-10800 S.C.C.P 🇧🇷 Amor Maior ♥
17	_bjdagorda		166	268	11765	124	-10800 Muita coisa eu mudaria, menos o fato de ser Gremista. 🇧🇷 GRÊMIO 🇧🇷 É p
18	_bolinhodoharry		494	526	12418	1363	-7200 Sim, meus ídolos são gays, próximo argumento?
19	_bubu_isa		97	105	2007	756	Facebook: Isadora Maciel, Insta: @isa_maciel, Whats: 51 82013722, Snap: isa.m
20	_cabuloosa		535	695	15818	1111	-10800
							⋮
2	Vertex	Followed	Followers	Tweets	Favorites	Time Zone UTC Offset (Seconds)	Description
3	_eviltwin		377	378	32826	3322	-10800 eu não assumo esse b.o.
4	_gabiimelo		648	1117	105052	638	-7200 17 anos, blogueira, futura jornalista, viciada em cafeína e um pouquinho hipst
5	_pamrodrigues		170	339	16670	834	-10800 Eu já briguei com a vida, hoje eu vivo bem com todo mundo...
6	_1marijuanna		850	809	7692	2073	-7200 Do what you want 🇧🇷 Você só aprendeu dar valor, quando eu aprendi a ficar sen
7	_aleceiffel		108	104	7451	1064	-10800 Little Eiffel.
8	_anotherdrink		424	706	43122	2545	-10800 o que me cita me excita
9	_atst		239	449	26954	8211	-10800 Nunca tem pizza no meu tomate
10	_aumenda		293	185	2791	376	-10800 Leonina ⚡, addicted TV series(vampire manic and Stelena), animals and futur
11	_belaseduz		3289	3567	13485	7085	-7200 Daora memo
12	_bellcristina		595	652	20606	7327	-10800 Seguindo nessa trilha íngreme e sem retorno tentando entender o que se pass
13	_bhannie		225	619	6997	3439	
14	_biaolive		1361	1536	4544	158	-7200 Mais amor
15	_bjdagorda		166	268	11837	130	-10800 Muita coisa eu mudaria, menos o fato de ser Gremista. 🇧🇷 GRÊMIO 🇧🇷 É p
16	_bolinhodoharry		507	535	12841	1398	-7200 Sim, meus ídolos são gays, próximo argumento?
17	_brendamarquess		337	366	51857	7035	-7200 you can't handle me
18	_brunaspeck		1729	881	8475	2524	-10800 paz, amor e muita sacanagem 🇧🇷   Snap : brunaspeck
19	_caamilamachado		453	1277	44555	1837	-10800
20	_cabrita_		242	969	60887	418	7200 Me ensina a crescer sem olhar pra trás e sentir medo #The4 #The5
							⋮

FONTE: O AUTOR (2015)

A Figura 28 mostra um recorte dos usuários que publicaram *tweets* relacionados à marca Avon, coletados durante as consultas realizadas no dia 06/04/2015 (parte superior da figura) e 13/04/2015 (parte inferior da figura). Pode-se perceber que os usuários “\_bjdagorda” e “\_bolinhodoharry” aparecem nas duas coletas, o que acarreta na duplicação desses usuários no momento de união dos arquivos.

Entretanto, diferente dos *tweets*, o NodeXL não oferece uma ferramenta própria para deduplicação dos perfis de usuários. Assim, fez-se necessário a utilização da funcionalidade “Remover Duplicatas”, disponibilizada pela aplicação de planilhas do *Office*. Essa função remove todas as linhas duplicadas com base nas colunas selecionadas, conforme ilustra a Figura 29.

**FIGURA 29 – EXECUÇÃO DA FUNÇÃO REMOVER DUPLICATAS, EXECUTADA PARA REMOÇÃO DOS USUÁRIOS DUPLICADOS**



FONTE: O AUTOR (2015)

A Tabela 5 coloca lado a lado o número de usuários obtidos na coleta e o número de usuários-únicos, obtidos após a deduplicação.

**TABELA 5 – COMPARATIVO ENTRE A QUANTIDADE DE USUÁRIOS ANTES E APÓS A DEDUPLICAÇÃO, REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS**

Marcas	Usuários	
	Antes	Após
Avon	27.618	22.048
Boticário	10.931	8.234
Jequiti	55.759	42.559
Natura	30.256	21.479
Total	124.564	94.320

FONTE: O AUTOR (2015)

Importante salientar dois pontos nesse processo. O primeiro deles é que após a deduplicação de usuários, os textos publicados pelos mesmos não são perdidos, uma vez que os *tweets* e usuários são armazenados em abas distintas da planilha. Assim, quando uma entrada referente a um usuário duplicado é eliminada da aba “Vertices”, o texto publicado pelo mesmo é mantido na aba “Edges”. Outro ponto é que durante a eliminação de usuários duplicados, sempre é mantido o registro da coleta mais recente. Mais uma vez tomando a Figura 28 como apoio, percebe-se que na coleta referente ao dia 06/04/2015, o perfil “\_bolinhodoharry” possuía 526 seguidores, contra 535 obtidos na coleta do dia 13/04/2015. Assim, após a deduplicação, o registro mantido foi aquele coletado no dia 13, obtendo-se, dessa forma, sempre os valores mais atualizados dos atributos.

#### 3.4.1.2.4 Atribuição de polaridade

O próximo passo executado para preparação dos dados foi a atribuição de polaridade aos *tweets* obtidos após a etapa de deduplicação, com a finalidade de determinar se a opinião que prevalece em relação ao *tweet* é positiva ou negativa.

Um grande desafio dessa etapa foi a escolha da abordagem a ser empregada, dentre aquelas apresentadas na subseção 2.3.4 dessa pesquisa, uma vez que a maioria dos dicionários léxicos e semânticos existentes (utilizados na maior parte das abordagens apresentadas), se aplicam à classificação de sentimentos para documentos na língua inglesa. Além disso, os dicionários léxicos encontrados na literatura para língua portuguesa foram construídos especificamente para o contexto em que seriam aplicados, conforme descrevem os trabalhos de mineração de opinião a partir de textos em língua portuguesa, apresentados na subseção 2.3.5.

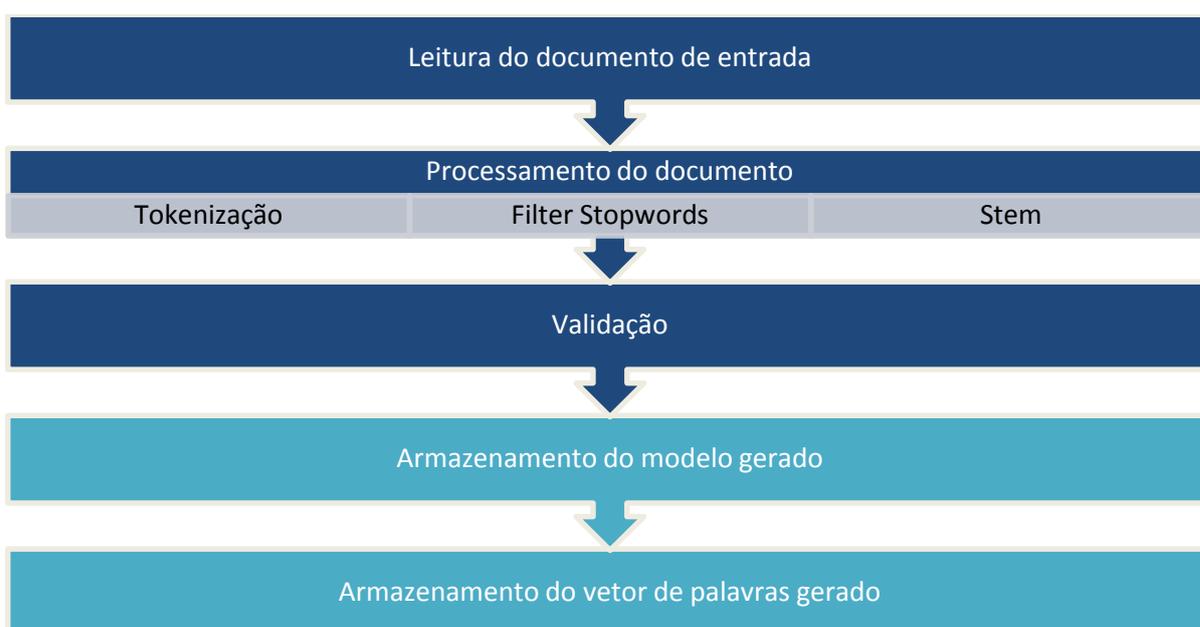
Assim, escolheu-se a abordagem baseada em aprendizado de máquina para a realização da tarefa de atribuição de polaridade aos *tweets*, uma vez que não necessita da utilização de dicionários léxicos ou semânticos, dado que um vetor de palavras é formado durante a fase de treinamento. Além disso, conforme descrito na subseção 2.3.5, abordagens baseadas em aprendizado de máquina, sobretudo aquelas que utilizam o algoritmo SVM, geralmente obtém boas taxas de acerto, em torno de 80%, considerando a precisão média da validação cruzada com três “*folds*”, utilizando uma base de teste pré-rotulada contendo *reviews* de filmes, sendo 700

rotulados previamente com sentimento positivo e outros 700 com sentimento negativo (TSYSARAU e PALPANAS, 2010).

O processo para atribuição de polaridade a textos utilizando aprendizagem de máquina é composto por duas fases, conforme apresentado na subseção 2.3.4. A etapa de treinamento, que pode ser supervisionada ou não supervisionada, e a etapa de classificação propriamente dita. Para implementação desse processo, foi utilizada a ferramenta *RapidMiner*.

A próxima decisão a ser tomada seria a escolha entre o aprendizado de máquina supervisionado e o aprendizado não supervisionado. Um processo de treinamento para cada uma das duas abordagens foi construído na ferramenta RapidMiner, representado por meio da Figura 30.

**FIGURA 30 – PASSOS EXECUTADOS DURANTE PROCESSO DE TREINAMENTO SUPERVISIONADO E NÃO-SUPERVISIONADO NA FASE DE PRÉ-TESTE**



FONTE: O AUTOR (2015)

Ambos os processos – supervisionado e não-supervisionado – diferenciam-se apenas pela forma como o arquivo de entrada foi construído, dado que no primeiro a rotulagem de sentimento é manual e no segundo, automatizada. Os testes para a escolha de qual das duas abordagens de aprendizado de máquina seria utilizado foram realizados com *tweets* da marca Avon. Assim, para o processo supervisionado, 501 *tweets* referentes a essa marca foram rotulados manualmente, como “positivos

ou negativos”; para o aprendizado não-supervisionado, foram coletados 22.255 *tweets* positivos e negativos, sobre qualquer tema na língua portuguesa, conforme apresentado na seção 3.4. Tanto o arquivo de entrada do processo supervisionado como do processo não-supervisionado estão no formato CSV, constituídos por dois atributos: *tweet* e sentimento associado.

A próxima etapa, nomeada “Processamento do documento”, realiza o processamento de texto propriamente dito, com base em três operações: *tokenize*, *filter stopwords* e *stem*. Essas operações são realizadas sobre cada *tweet*.

O operador “*Tokenize*” divide o texto do *tweet* em “pedaços”, em que cada pedaço consiste em uma única palavra; o operador “*Filter Stopwords*”, por sua vez, remove palavras tidas como irrelevantes para a etapa de classificação, como artigos (definidos, indefinidos e suas combinações), pronomes e até mesmo alguns verbos que não denotam sentimento. Para finalizar a etapa de processamento de documento, é realizada a operação de “*Stem*”, que mantém apenas o radical das palavras; assim as palavras “gostei”, “gostaram”, “gostoso” se transformam todas em “gost” após a aplicação do operador.

A próxima etapa executada é a de validação, dividida em duas subetapas: treinamento e teste. De modo que a validação possa ser realizada, cada uma das bases de registros a serem classificados (formada por 501 registros no caso da base supervisionada e 22.255 no caso da base não supervisionada) é dividida em dois “*splits*” (partições), sendo que a partição maior, configurada para ser formada por 70% da base, foi utilizada para o treinamento e os 30% restantes, para a fase de testes.

A fase de treinamento aplicou o algoritmo SVM em sua implementação JMySVM para determinar como cada uma das palavras que compõe o vetor de palavras – obtido na fase de processamento de documentos – influencia para que o sentimento do *tweet* seja positivo ou negativo. O modelo gerado pelo algoritmo SVM foi então aplicado ao restante da base e o desempenho é calculado comparando-se o valor rotulado de cada *tweet* e o valor inferido pelo modelo gerado pelo algoritmo SVM.

A precisão alcançada no processo de aprendizado supervisionado foi de 77,33%, conforme detalhado na matriz de confusão representada pela Tabela 6.

**TABELA 6 – MATRIZ DE CONFUSÃO OBTIDA COMO SAÍDA DO PROCESSO DE TREINAMENTO SUPERVISIONADO COM EMPREGO DO ALGORITMO SVM DURANTE A FASE DE PRÉ-TESTE**

	Verdadeiro POSITIVO	Verdadeiro NEGATIVO	Precisão da classe
Predição POSITIVO	98	30	76,56%
Predição NEGATIVO	4	18	81,82%
Cobertura da classe	96,08%	37,50%	
			Precisão: 77,33%

**FONTE: O AUTOR (2015)**

Para o processo de aprendizado não supervisionado, a precisão obtida foi de 80,28%, conforme demonstra a matriz de confusão representada por meio da Tabela 7.

**TABELA 7 – MATRIZ DE CONFUSÃO OBTIDA COMO SAÍDA DO PROCESSO DE TREINAMENTO NÃO-SUPERVISIONADO COM EMPREGO DO ALGORITMO SVM DURANTE A FASE DE PRÉ-TESTE**

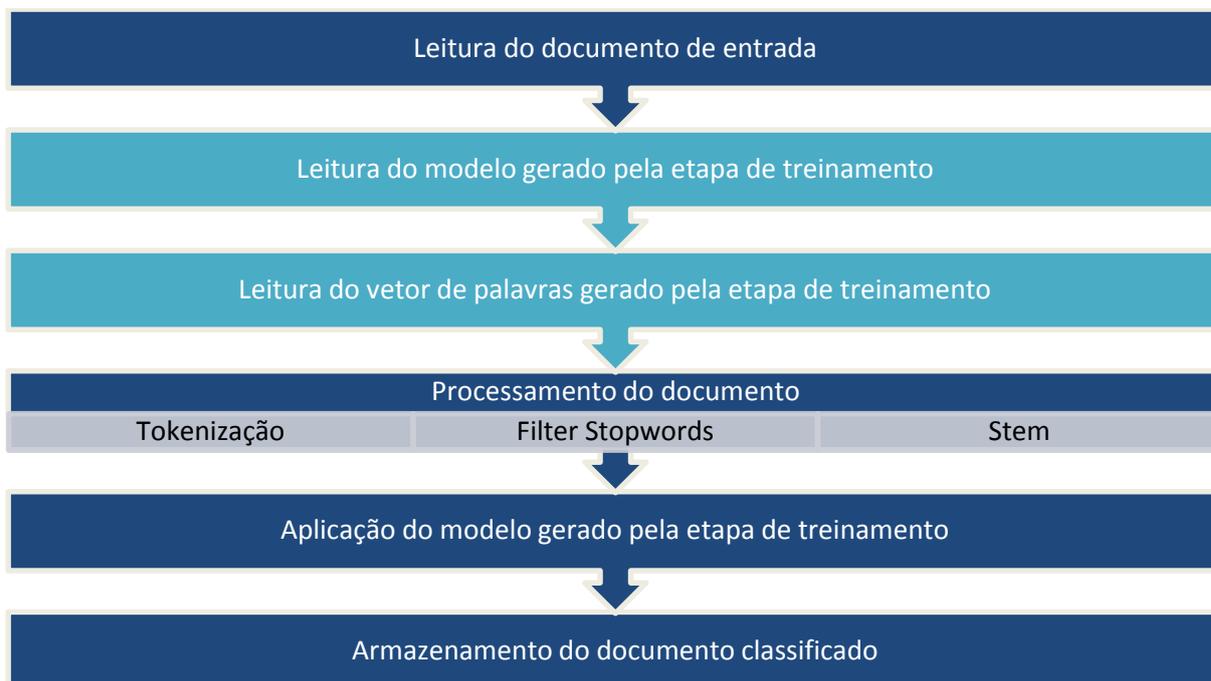
	Verdadeiro POSITIVO	Verdadeiro NEGATIVO	Precisão da classe
Predição POSITIVO	2729	385	87,64%
Predição NEGATIVO	907	2530	73,61%
Cobertura da classe	75,06%	86,79%	
			Precisão: 80,28%

**FONTE: O AUTOR (2015)**

A partir dos resultados observados na Tabelas 6 e na Tabela 7, decidiu-se, para a atribuição de polaridade de sentimento aos 148.050 *tweets* que foram analisados nessa pesquisa, utilizar-se o modelo gerado pelo processo de treinamento não-supervisionado, pelos seguintes motivos: a) a precisão alcançada no processo de treinamento não-supervisionado é satisfatória e ligeiramente superior à precisão obtida por meio do processo supervisionado; b) é característica intrínseca do processo não-supervisionado a não necessidade de rotulagem manual dos *tweets*; assim, não se demanda tempo para a rotulagem destes; c) a base de *tweets* utilizada no aprendizado supervisionado é independente de contexto e pode, portanto, ser aplicada a qualquer conjunto de *tweets* em português que precise ser classificado.

Ao final do processo de treinamento, são geradas duas saídas: o vetor de palavras e o modelo gerado pelo algoritmo SVM. Essas duas saídas serão utilizadas como entradas para o processo de classificação propriamente dita, representado pela Figura 32.

FIGURA 32 – PASSOS EXECUTADOS DURANTE PROCESSO DE TREINAMENTO SUPERVISIONADO E NÃO-SUPERVISIONADO NA FASE DE PRÉ-TESTE



FONTE: O AUTOR (2015)

O processo de classificação dos *tweets* inicia-se com a leitura de um arquivo CSV contendo os *tweets* a serem classificados. Ao todo, são quatro arquivos, referentes aos *tweets* de cada uma das quatro marcas utilizadas no pré-teste.

A leitura dos artefatos gerados na etapa de treinamento – modelo e vetor de palavras – é realizada. O vetor de palavras é utilizado como uma das entradas da etapa de “Processamento do documento”; a outra entrada é o documento em si, contendo os *tweets* a serem classificados. O processamento de texto realizado nessa etapa é idêntico ao realizado no processo de treinamento, em que cada *tweet* é submetido sequencialmente aos operadores “*Tokenize*”, “*Filter Stopwords*” e “*Stem*”. A diferença do processo de treinamento para o processo de classificação é que o vetor de palavras obtido como saída deste contém palavras que aparecem nos *tweets* (após as três operações de processamento de texto descritas anteriormente) e que, ao mesmo tempo, constem no vetor de palavras gerado pelo processo de treinamento.

A etapa seguinte do processo de classificação, “*Aplicação do modelo gerado pela etapa de treinamento*”, recebe como entrada os *tweets* vetorizados (saída da etapa “*Processamento do documento*”) e o modelo gerado pelo processo de treinamento. A partir dessas duas entradas, o operador classifica cada um dos *tweets*

como “positivo” ou “negativo”. Por fim, o documento contendo os *tweet* se suas respectivas polaridades é armazenado.

Assim, após a execução do processo de classificação pela quarta vez, cada um dos 148.050 *tweets* coletados referentes às marcas Avon, Boticário, Jequití e Natura estão classificados quanto ao sentimento predominante.

#### 3.4.1.2.5 *Obtenção das métricas para análise de redes sociais*

O passo que antecede a aplicação do modelo proposto para avaliação do *brand equity* das quatro marcas utilizadas no pré-teste é a obtenção das métricas para análise de cada uma das redes construídas pelo processo de coleta de *tweets*. Conforme apresentado na subseção 2.2.5, as métricas de ARS podem ser úteis para identificar propriedades da rede como um todo, bem como de cada um de seus vértices.

A ferramenta NodeXL realiza esses cálculos com base nos vértices e arestas recuperados a partir da busca de *tweets*, tendo como base a estrutura descrita nos Quadros 3 e 4. O cálculo das métricas é realizado por meio da funcionalidade “*Graph Metrics*”, que executa o cálculo de métricas do grafo – número de vértices, número de arestas únicas, número de arestas duplicadas, número total de arestas, número de auto loops, número de componentes, número que componentes que possuem um único vértice, número máximo de vértices em um componente conectado, número máximo de arestas em um componente conectado, máxima distância geodésica (diâmetro), distância geodésica média e a densidade do grafo – e referentes a cada uma das arestas – grau de entrada, grau de saída, grau de centralidade, centralidade de intermediação e centralidade de proximidade.

Assim, ao final das etapas de coleta e preparação dos dados, obteve-se a seguinte estrutura: a) *tweets* de cada uma das quatro marcas do pré-teste organizados na estrutura de uma rede social – com as postagens constituindo as arestas e os perfis envolvidos constituindo os vértices dessa rede; b) polaridade positiva ou negativa atribuída a cada um dos *tweets* descritos em (a); c) métricas da rede como um todo e específicas de cada vértice calculadas a partir da estrutura descrita em (a); d) identificador dos seguidores (*followers*) dos perfis referentes a cada uma das marcas do pré-teste.

A partir desses dados, foi possível aplicar o modelo preliminar proposto para avaliação do *brand equity* de cada uma das marcas.

### 3.4.2 Teste

A etapa de teste foi executada com a finalidade de verificar o impacto das melhorias propostas com base na análise dos resultados da aplicação do modelo preliminar resultante da etapa de pré-teste.

#### 3.4.2.1 Coleta de Dados II

Na etapa de teste, o modelo modificado foi aplicado sobre um conjunto de *tweets* diferente daquele utilizado na etapa de pré-teste, com a finalidade de demonstrar que o modelo é independente de contexto, ou seja, pode ser aplicado sobre conjuntos de marcas de diferentes categorias de produtos.

Assim, ao invés de a análise ser referente a marcas de cosméticos, foram explorados *tweets* concernentes a marcas de montadoras de automóveis. As marcas selecionadas para o estudo foram Chevrolet ([twitter.com/chevroletbrasil](https://twitter.com/chevroletbrasil)), Fiat ([twitter.com/fiatbr](https://twitter.com/fiatbr)), Renault ([twitter.com/renaultbrasil](https://twitter.com/renaultbrasil)) e Volkswagen ([twitter.com/carromesmo](https://twitter.com/carromesmo)). Dois critérios foram adotados para seleção das marcas: a) as marcas deveriam estar entre as que mais venderam carros no Brasil em 2014, de acordo com levantamento feito pela revista Exame (BARBOSA, 2014); b) as marcas deveriam possuir perfil oficial em português no Twitter.

Foram realizadas cinco coletas de *tweets* relacionados às quatro marcas selecionadas, entre os dias 06/07/2015 e 03/08/2015. Assim como na fase de pré-teste, as extrações foram realizadas com o auxílio da ferramenta NodeXL; como a janela permitida para busca é de, aproximadamente, sete dias, as coletas resultaram em *tweets* escritos entre os dias 27/06/2015 e 03/08/2015.

Nesse período, foram publicados 76.003 *tweets* referentes às quatro montadoras, formando uma rede de relacionamento composta por 49.450 usuários, conforme mostra a Tabela 8.

**TABELA 8 – CONSULTAS REALIZADAS ENTRE 06/07/2015 E 03/08/2015 E NÚMERO DE TWEETS E USUÁRIOS RETORNADOS PARA CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS PESQUISADAS**

Marcas	Consulta	Tweets	Usuários
Chevrolet	"chevrolet lang:pt"	14.818	8.387
Fiat	"fiat lang:pt"	24.806	17.881
Renault	"renault lang:pt"	11.933	7.742
Volkswagen	"volkswagen lang:pt"	24.446	15.440
Total	N/A	76.003	49.450

FONTE: O AUTOR (2015)

Assim como na etapa de pré-teste, todos os atributos disponibilizados pela ferramenta NodeXL foram coletados, resultando em um total de 13 atributos referentes aos *tweets* (*relationship*, *vertex1*, *vertex2*, *relationship date*, *tweet*, *urls in tweet*, *domains in tweet*, *hashtags in tweet*, *tweet date*, *twitter page for tweet*, *latitude*, *longitude*, *imported id*, *in-reply-to tweet id*) e mais 16 atributos referentes aos usuários (*vertex*, *image file*, *followed*, *followers*, *tweets*, *favorites*, *time zone*, *utc offset*, *description*, *location*, *web*, *time zone*, *joined twitter date*, *custom menu*, *item text*, *tweeted search item*), detalhados previamente na seção 3.3.1.

Os *tweets* que formaram o conjunto de treinamento para a etapa de classificação de sentimentos também foram os mesmos do pré-teste, já que os mesmos são independentes de contexto.

Para que o modelo proposto pudesse ser aplicado, foi necessário ainda coletar os identificadores dos seguidores de cada um dos perfis representantes das quatro montadoras selecionadas. Essa coleta foi realizada em 20/08/2015 e, portanto, representa os seguidores dos perfis naquele momento. A Tabela 9 mostra o número total de identificadores coletados de cada uma das quatro marcas.

**TABELA 9 – NÚMERO DE SEGUIDORES OBTIDOS A PARTIR DA COLETA DE CADA UM DOS PERFIS ASSOCIADOS ÀS MARCAS AUTOMOBILÍSTICAS ANALISADAS**

Marcas	Perfil	Seguidores
Chevrolet	twitter.com/chevroletbrasil	57.187
Fiat	twitter.com/fiatbr	92.281
Renault	twitter.com/renaultbrasil	83.380
Volkswagen	twitter.com/carromesmo	42.034
Total	N/A	274.882

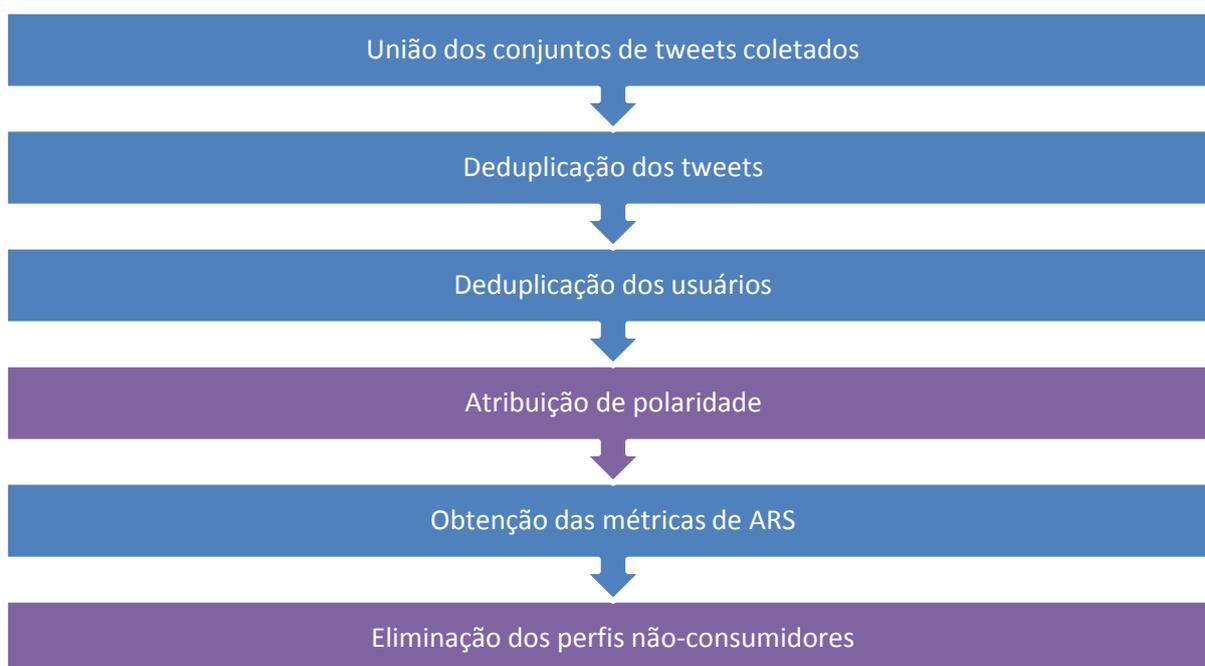
FONTE: O AUTOR (2015)

Finalizada a coleta dos *tweets* dos seguidores dos perfis relacionados a cada uma das quatro marcas selecionadas para o teste, os dados foram preparados para ser utilizados como entrada para o novo modelo proposto. Essa preparação é descrita na seção 3.4.2.2.

### 3.4.2.2 *Preparação dos Dados II*

A etapa de preparação dos dados para o modelo proposto sofreu também algumas alterações em relação à preparação dos dados do modelo preliminar, decorrentes das diferenças entre os dois modelos. Os passos dessa preparação são mostrados na Figura 33, sendo destacados os passos alterados ou incluídos.

**FIGURA 33 – PASSOS EXECUTADOS NO TESTE DURANTE A ETAPA DE PREPARAÇÃO DOS DADOS**



**FONTE: O AUTOR (2015)**

A execução de cada um dos passos é apresentada nas subseções a seguir.

#### 3.4.2.2.1 *União dos conjuntos de tweets coletados*

As cinco coletas dos *tweets* para a segunda fase da pesquisa, também realizadas por meio da ferramenta NodeXL, retornaram um total de 20 arquivos

referentes a cada uma das marcas consultadas – Chevrolet, Fiat, Renault e Volkswagen. Foi criado então um novo arquivo para cada uma das marcas, que passou a conter todos os *tweets* recuperados referentes àquela marca.

#### 3.4.2.2.2 *Deduplicação dos tweets*

Mais uma vez, a etapa de união dos conjuntos de *tweets* coletados implicou na duplicação de alguns *tweets*, causada pela sobreposição das janelas de buscas. Assim, foram novamente aplicados os procedimentos para deduplicação utilizados na etapa de pré-teste. O comparativo entre o número de *tweets* antes e após a etapa de deduplicação é mostrado na Tabela 10.

**TABELA 10 – COMPARATIVO ENTRE A QUANTIDADE DE TWEETS ANTES E APÓS A DEDUPLICAÇÃO, REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS ANALISADAS**

Marcas	Tweets	
	Antes	Após
Chevrolet	14.818	12.610
Fiat	24.806	21.181
Renault	11.933	9.411
Volkswagen	24.446	21.471
Total	76.003	64.673

**FONTE: O AUTOR (2015)**

Como se pode constatar a partir dos dados apresentados na Tabela 10, o processo de deduplicação de *tweets* resultou na redução de 14,9% da amostragem inicial.

#### 3.4.2.2.3 *Deduplicação dos usuários*

Da mesma forma, o processo de deduplicação de usuários não sofreu alterações em relação ao processo aplicado na etapa de pré-teste. A Tabela 11 apresenta a quantidade de usuários coletados, por marca, antes e após o processo de deduplicação de usuários.

**TABELA 11 – COMPARATIVO ENTRE A QUANTIDADE DE USUÁRIOS ANTES E APÓS A DEDUPLICAÇÃO, REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS ANALISADAS**

Marcas	Usuários	
	Antes	Após
Chevrolet	8.387	5.736
Fiat	17.881	12.867
Renault	7.742	5.073
Volkswagen	15.440	11.349
Total	49.450	35.025

**FONTE: O AUTOR (2015)**

A partir da verificação do número total de usuários antes e após o processo de deduplicação, pode-se constatar uma diminuição de 29,2% em relação à coleta inicial.

#### 3.4.2.2.4 *Atribuição de polaridade*

O processo de atribuição de polaridade – positiva ou negativa – a cada um dos *tweets* resultantes dos processos de deduplicação sofreu poucas alterações em relação aos procedimentos empregados na etapa de pré-teste.

Novamente, a abordagem baseada em aprendizado de máquina não-supervisionada foi empregada, dividida mais uma vez em duas etapas: treinamento e classificação. Durante a etapa de treinamento, utilizou-se a mesma base não-supervisionada da etapa de pré-teste, formada por *tweets* previamente rotulados como positivos ou negativos. Uma vez que essa base é independente de contexto, a mesma pôde ser empregada tanto para a geração do modelo utilizado para classificar *tweets* referentes às marcas de cosméticos como para classificar *tweets* relativos às montadoras de veículos.

A etapa de treinamento, entretanto, sofreu uma alteração. O tipo de amostragem utilizada passou a ser a estratificada, ao contrário da amostragem aleatória utilizada no pré-teste. Assim, garantiu-se que os subconjuntos utilizados no processo de validação possuem a mesma distribuição de classes presentes na amostra completa. A taxa de *split* foi a mesma utilizada no pré-teste, ou seja, 0,7. Assim, 70% da base foram utilizadas para o treinamento e os 30% restantes, para validação. A matriz de confusão resultante do processo de treinamento não-

supervisionado com emprego do algoritmo SVM em sua implementação JMySVM, taxa de *split* de 0,7 e amostragem estratificada pode ser verificada na Tabela 12.

**TABELA 12 – MATRIZ DE CONFUSÃO OBTIDA COMO SAÍDA DO PROCESSO DE TREINAMENTO SUPERVISIONADO COM EMPREGO DO ALGORITMO SVM, DURANTE A FASE DE TESTE**

	Verdadeiro POSITIVO	Verdadeiro NEGATIVO	Precisão da classe
Predição POSITIVO	2738	372	88,04%
Predição NEGATIVO	863	2578	74,92%
Cobertura da classe	75,06%	86,79%	
			Precisão: 81,15%

**FONTE: O AUTOR (2015)**

Com as alterações empregadas na seleção da amostragem, pode-se observar uma melhoria na taxa de acerto, que passou de 80,28% (etapa de pré-teste) para 81,15% (etapa de teste).

Após o processo de treinamento, executou-se o processo de classificação dos *tweets* referentes às quatro marcas de montadoras, seguindo os mesmos passos apresentados anteriormente na Figura 32. Finalizado esse processo, os 64.673 *tweets* resultantes do processo de deduplicação estavam classificados quanto ao sentimento predominante.

#### 3.4.2.2.5 *Obtenção das métricas de ARS*

Assim como na etapa de pré-teste, utilizou-se a ferramenta NodeXL para obtenção das métricas utilizadas na análise de redes sociais, com destaque para a métrica PageRank, que foi utilizada nos modelos propostos para mensurar a importância de cada elemento presente na rede de usuários, a partir das características de suas ligações com os demais elementos da rede.

#### 3.4.2.2.6 *Eliminação dos perfis não-consumidores*

O processo de eliminação dos perfis não-consumidores não foi empregado durante a etapa de pré-teste. Contudo, observou-se que os perfis dos usuários relativos a empresas, revendedores, *spammers*, mídias jornalísticas, personagens fictícios e perfis de teor humorístico exerciam grande influência sobre alguns

elementos do modelo preliminar. Uma vez que o modelo proposto busca compreender como os consumidores percebem as marcas, é inconveniente incluir nessa análise a percepção de perfis que não sejam, potencialmente, consumidores.

Assim, para tentar diminuir o impacto dos perfis não-consumidores na análise, buscou-se classificar os perfis coletados e, com base nessa classificação, manter apenas os *tweets* publicados por perfis classificados como consumidores. Para isso, os perfis dos usuários recuperados durante a etapa de coleta foram classificados em “Consumidores” e “Não-Consumidores”. Empregou-se, para essa finalidade, o algoritmo de classificação C4.5, em sua implementação J48, utilizando-se a ferramenta Weka.

Inicialmente, para construção do modelo de classificação, foram utilizados os atributos referentes aos usuários, obtidos durante a etapa de coleta dos dados. Esses atributos são: *vertex* (nome do perfil do usuário no Twitter), *followed* (número de perfis seguidos pelo usuário), *followers* (número de seguidores do usuário), *tweets* (número de *tweets* publicados pelo usuário), *favorites* (número de *tweets* marcados como favorito pelo usuário) e *joined twitter date* (data em que se tornou usuário da rede social Twitter).

Para minimizar os problemas resultantes das dispersões distintas entre essas variáveis durante o processo de construção da árvore de classificação, os atributos *followed*, *followers*, *tweets* e *favorites* foram normalizados de modo que seus valores ficassem entre 0 e 1.

Adicionalmente, a partir dos atributos *joined twitter date* e *tweets*, extraiu-se um novo atributo, utilizado para representar a média de *tweets* publicados por dia. Esse novo atributo, denominado “*tweets por dia*” também foi normalizado, de modo que seus valores ficassem na faixa entre 0 e 1.

Um último atributo, denominado “nome”, foi criado a partir do cruzamento do atributo *vertex* com uma base de nomes brasileiros. Para a construção dessa base, foram utilizadas as listas de aprovados nos vestibulares da Fuvest entre os anos de 2011 e 2015. Após a extração do primeiro nome de cada um dos aprovados, a deduplicação de nomes e a eliminação de nomes com quatro caracteres ou menos, uma lista final com 6.753 nomes foi obtida.

A lista de aprovados da Fuvest foi escolhida como ponto de partida para construção dessa base, dada a grande variedade de nomes (maior do que em outros

dicionários de nomes testados) e também pela disponibilidade dessas listas em formato .txt, que apresenta uma maior facilidade de manipulação.

Assim, caso um item presente na base de nomes fosse encontrado em parte do nome do perfil do usuário no Twitter, seria atribuído “sim” ao valor da variável nome ou “não”, caso contrário. Dessa forma, seria atribuído “sim” para o atributo nome referente ao perfil @eduardo UFPR e o valor “não” se o perfil for, por exemplo, @UFPR. Nomes contendo quatro caracteres ou menos foram eliminados da base de nomes para evitar falsos positivos. Se, por exemplo, o nome “Ana” fosse mantido na base de nomes, seria atribuído “sim” à variável nome correspondente ao perfil @banana\_nanica, o que estaria incorreto.

Por fim, criou-se um atributo meta que poderia receber os valores “CONSUMIDOR” ou “NÃO-CONSUMIDOR”. Uma base contendo 200 perfis foi manualmente classificada para que pudesse ser utilizada como entrada para a geração da árvore de classificação. Dos 200 perfis selecionados para compor a base de treinamento, 100 foram classificados como consumidores e 100 como não consumidores. Parte dessa base pode ser observada no Quadro 5.

**QUADRO 5 – AMOSTRA PARCIAL DA BASE DE USUÁRIOS UTILIZADA NO PROCESSO DE CLASSIFICAÇÃO DOS PERFIS EM DUAS CATEGORIAS: CONSUMIDORES E NÃO-CONSUMIDORES**

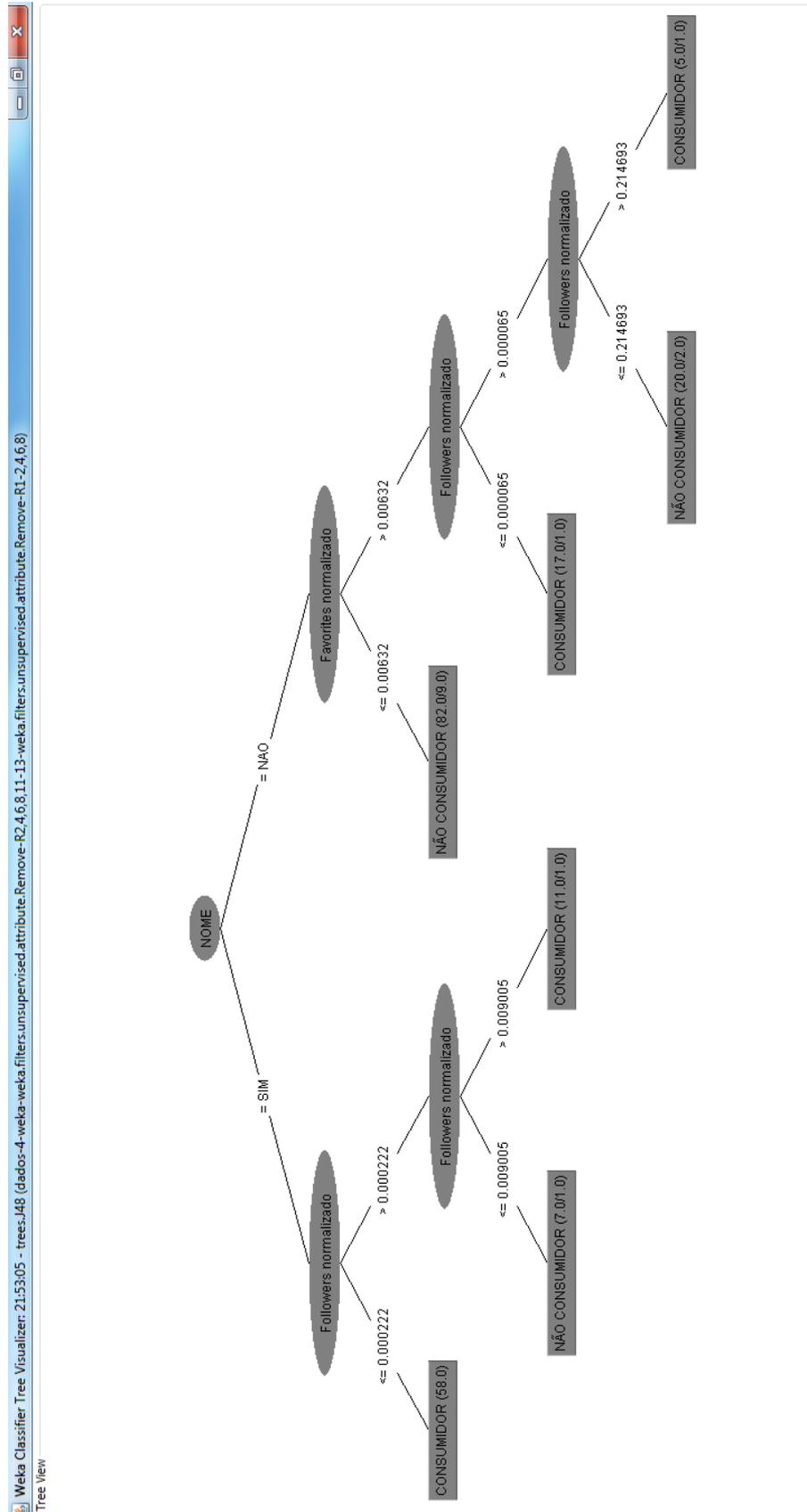
Vertex	Followed normalizado	Followers normalizado	Tweets normalizado	Favorites normalizado	Joined Twitter	tweets/dia normalizado	NOME	CLASSE
_rodrigoneves	0,0009	0,0001	0,0332	0,0005	03/08/09	0,0170	SIM	CONSUMIDOR
adrianacardoso	0,0003	0,0000	0,0321	0,0165	14/05/10	0,0190	SIM	CONSUMIDOR
adrianadanielis	0,0015	0,0000	0,0152	0,1292	28/04/13	0,0222	SIM	CONSUMIDOR
adynefigueired	0,0002	0,0000	0,0150	0,0070	15/04/11	0,0110	NAO	CONSUMIDOR
agazetadoacre	0,0004	0,0000	0,0348	0,0000	24/02/10	0,0198	NAO	NÃO CONSUMIDOR
aillavitoriana	0,0007	0,0000	0,0085	0,0395	03/11/11	0,0072	SIM	CONSUMIDOR
alanguerra43	0,0068	0,0000	0,0253	0,0081	28/02/15	0,4197	NAO	CONSUMIDOR
alinedelima	0,0003	0,0000	0,0083	0,0016	06/06/09	0,0042	SIM	CONSUMIDOR
amandakamins	0,0005	0,0000	0,0672	0,2922	26/11/10	0,0447	SIM	CONSUMIDOR
amandaschimid	0,0002	0,0000	0,0013	0,0134	15/12/14	0,0103	SIM	CONSUMIDOR
anabeatrizvaz_	0,0008	0,0000	0,0337	0,0705	08/01/12	0,0300	SIM	CONSUMIDOR
annabeatriz_lin	0,0003	0,0000	0,0023	0,0084	21/09/11	0,0019	SIM	CONSUMIDOR
avonbr	0,0058	0,0014	0,0525	0,0000	09/03/09	0,0252	NAO	NÃO CONSUMIDOR
avonmaquiager	0,0972	0,0015	0,0104	0,0002	09/06/09	0,0052	NAO	NÃO CONSUMIDOR
bandsportstv	0,0010	0,0017	0,1032	0,0002	17/11/09	0,0557	NAO	NÃO CONSUMIDOR
barbiesemken	0,4976	0,0116	0,1457	0,0121	20/07/09	0,0743	NAO	NÃO CONSUMIDOR
beatrizmilao	0,0005	0,0000	0,0090	0,0127	09/05/13	0,0135	SIM	CONSUMIDOR
blog_nostillo	0,0049	0,0000	0,0100	0,0007	02/07/11	0,0077	NAO	NÃO CONSUMIDOR
blogdoecomme	0,1978	0,0016	0,1740	0,0000	15/10/09	0,0925	NAO	NÃO CONSUMIDOR
bloglovin	0,0009	0,0008	0,0084	0,0433	24/09/08	0,0038	NAO	NÃO CONSUMIDOR
blogsbelleza	0,0389	0,0005	0,3176	0,0004	17/05/13	0,4775	SIM	NÃO CONSUMIDOR
bomsenhora	0,0003	0,0004	0,0963	0,6989	09/06/09	0,0481	NAO	NÃO CONSUMIDOR

FONTE: O AUTOR (2015)

Após etapa de preparação a funcionalidade do Weka denominada “*select attributes*” foi executada para identificar os principais atributos que exerciam influência sobre o atributo meta. Nessa etapa, os atributos selecionados foram: “*followers normalizado*”, “*favorites normalizado*”, “*tweets/dia normalizado*” e “nome”.

Assim, a aplicação do algoritmo de classificação J48 sobre o conjunto de dados resultou na árvore mostrada na Figura 34, com uma taxa de acerto de 86,5%, utilizando-se validação cruzada de dez partições.

FIGURA 34 – ÁRVORE GERADA PELO ALGORITMO C4.5 (EM SUA IMPLEMENTAÇÃO J48) PARA CLASSIFICAÇÃO DOS PERFIS DO TWITTER EM DUAS CATEGORIAS: CONSUMIDORES E NÃO-CONSUMIDORES



FONTE: O AUTOR (2015)

Assim, a árvore gerada pelo algoritmo de classificação foi utilizada para eliminação dos perfis não-consumidores da base, bem como dos *tweets* por eles publicados. O impacto dessa etapa pode ser visualizado na Tabela 13 abaixo.

**TABELA 13 – COMPARATIVO ENTRE O NÚMERO DE USUÁRIOS E TWEETS ANTES E APÓS A ELIMINAÇÃO DOS PERFIS CLASSIFICADOS COMO NÃO-CONSUMIDORES, AGRUPADOS PELAS MARCAS AUTOMOBILÍSTICAS ANALISADAS**

Marcas	Usuários			Tweets		
	Antes	Após	%	Antes	Após	%
<b>Chevrolet</b>	5.736	2.653	-53,7%	12.610	3.699	-70,7%
<b>Fiat</b>	12.867	7.540	-41,4%	21.181	10.387	-51,0%
<b>Renault</b>	5.073	2.596	-48,8%	9.411	3.796	-59,7%
<b>Volkswagen</b>	11.349	5.700	-49,8%	21.471	8.190	-61,9%
<b>Total</b>	<b>35.025</b>	<b>18.489</b>	<b>-47,2%</b>	<b>64.673</b>	<b>26.072</b>	<b>-59,7%</b>

FONTE: O AUTOR (2015)

A partir da análise dos números apresentados na Tabela 13, observou-se a diminuição de 47,2% dos usuários presentes na base. Esses usuários eram responsáveis pela publicação de 59,7% dos *tweets*, os quais foram igualmente eliminados.



Nas subseções 4.1.1, 4.1.2, 4.1.3 e 4.1.4 são apresentadas as propostas para análise de cada um dos quatro elementos a partir dos dados obtidos das redes sociais.

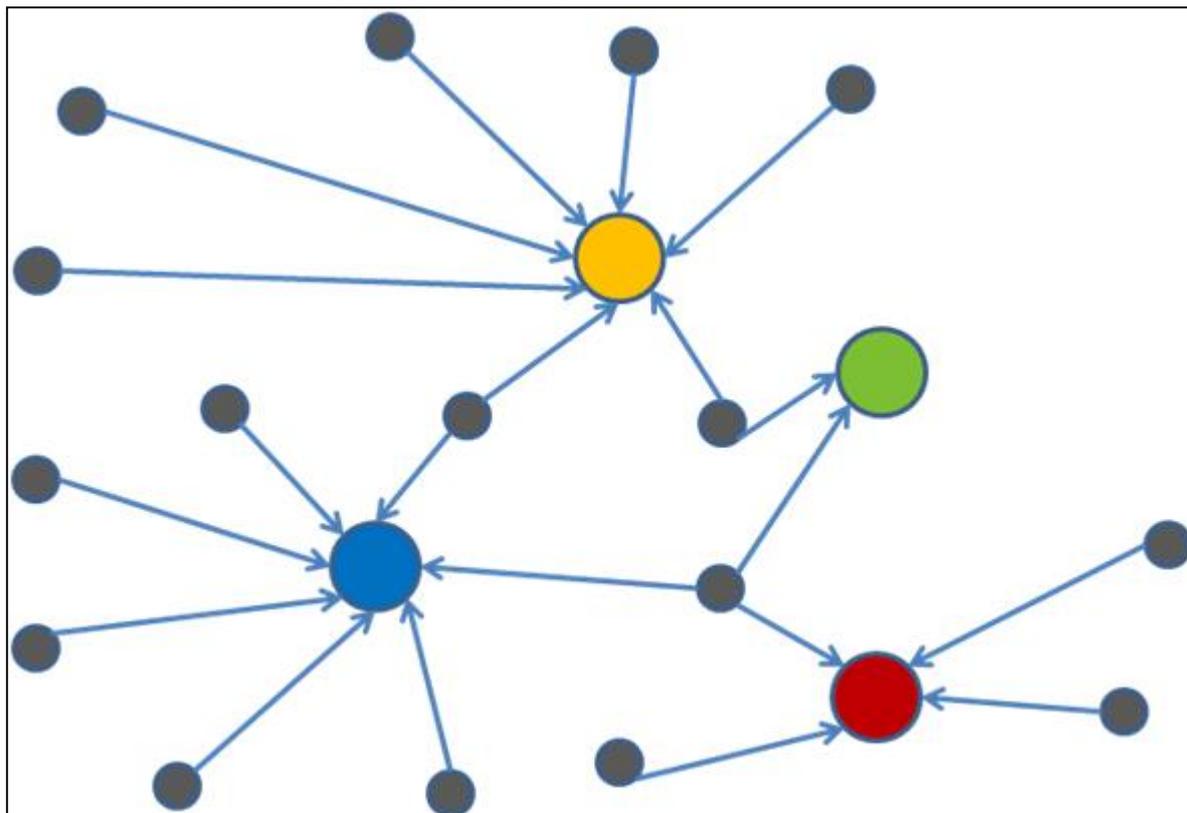
#### 4.1.1 Conhecimento da Marca

Conforme Aaker define, “o conhecimento da marca é a capacidade que um comprador potencial tem de reconhecer ou se recordar de uma marca como integrante de uma categoria de produtos” (AAKER, 1998, p. 64) ou ainda a força da presença da marca na mente dos consumidores (AAKER, 1998; KELLER, 1993).

Com base nessas definições, tanto Yoo e Donthu (2001) quanto Pappu, Quester e Cooksey (2005) elaboraram questionários com o objetivo de medir o conhecimento dos consumidores em relação à determinada marca, aplicando perguntas como “Você conhece a marca X?” ou “Qual a primeira marca de refrigerantes que vêm a sua mente?”

Nas redes sociais, propõe-se que o conhecimento da marca possa ser expresso a partir de uma relação direta entre o número de seguidores de uma marca e o conhecimento dessa marca na rede social. Assim, uma rede de seguidores das marcas de uma determinada categoria de produtos, pode ser representada pelo grafo direcionado  $G = (V, E)$ , no qual cada vértice representa um perfil na rede social Twitter e suas arestas representam a relação “seguir”, de modo que a ligação formada pelo par de vértices  $\{u, v\}$  significa que “u segue v” ou que “v é seguido por u”. A Figura 36 é apresentada para facilitar a compreensão dessa proposta.

FIGURA 36 – PROPOSTA DO MODELO PREMILINAR PARA ESTABELECIMENTO DO CONHECIMENTO DA MARCA A PARTIR DE UMA REDE DE SEGUIDORES



FONTE: O AUTOR (2015)

Na Figura 36, os vértices coloridos (amarelo, verde, azul e vermelho) representam os perfis de quatro marcas, enquanto que os vértices mais escuros representam seus respectivos seguidores. Assim, as marcas representadas pelos vértices amarelo e azul são as que possuem maior conhecimento nessa rede, com sete seguidores cada, enquanto que a marca representada pelo vértice azul é a menos conhecida na rede com apenas dois seguidores. Deste modo, o conhecimento de uma marca na rede social pode ser expresso pelo grau de entrada do vértice que representa a marca na rede de seguidores.

Logo:

$$c(x) = grau\_entrada(x) \quad (18)$$

Onde:

$c(x)$ : conhecimento da marca  $x$

De modo análogo, o conhecimento normalizado de uma marca  $x$  poderia ser dado por:

$$c'(x) = \frac{c(x)}{n} \quad (19)$$

Onde:

$c'(x)$ : conhecimento normalizado da marca  $x$

$c(x)$ : conhecimento da marca  $x$

$n$ : número total de vértices que formam a rede de seguidores

Logo, quanto maior o número de seguidores de uma marca, maior o conhecimento dessa marca na rede social.

#### 4.1.2 Lealdade à Marca

Relembrando o conceito proposto por Aaker, a lealdade a uma marca pode ser medida por meio da ligação do consumidor com a marca, de modo que quanto maior for a lealdade do consumidor àquela marca, menor a chance de trocá-la, nem mesmo ocasionalmente, por uma marca da concorrência. Nos modelos tradicionais, a lealdade pode ser medida de forma direta por meio dos padrões reais de compra, sugerindo três medições possíveis: taxas de recompra, percentual de compras e número de marcas compradas (AAKER, 1998).

Os modelos operacionais propostos por Yoo e Donthu (2001) e Pappu, Quester e Cooksey (2005) para medição da lealdade do consumidor em relação a uma marca também partem da definição proposta por Aaker (1998); outra definição aplicada por esses modelos é a de Oliver (2014), segundo a qual a lealdade à marca está relacionada ao compromisso consistente de recompra de um produto ou serviço no futuro, mesmo que submetidos a influências e esforços de marketing de marcas concorrentes. Assim, para tentar mensurar a lealdade do consumidor à marca, esses trabalhos aplicaram questionários aos consumidores que continham perguntas como “Você se considera leal à marca X?”, “A marca X seria sua primeira escolha?” ou ainda “Você compraria outras marcas se a marca X estivesse disponível na loja?”

Desta forma, a lealdade do consumidor em relação a uma determinada marca parece sugerir uma relação de exclusividade, representada pelos níveis mais altos da

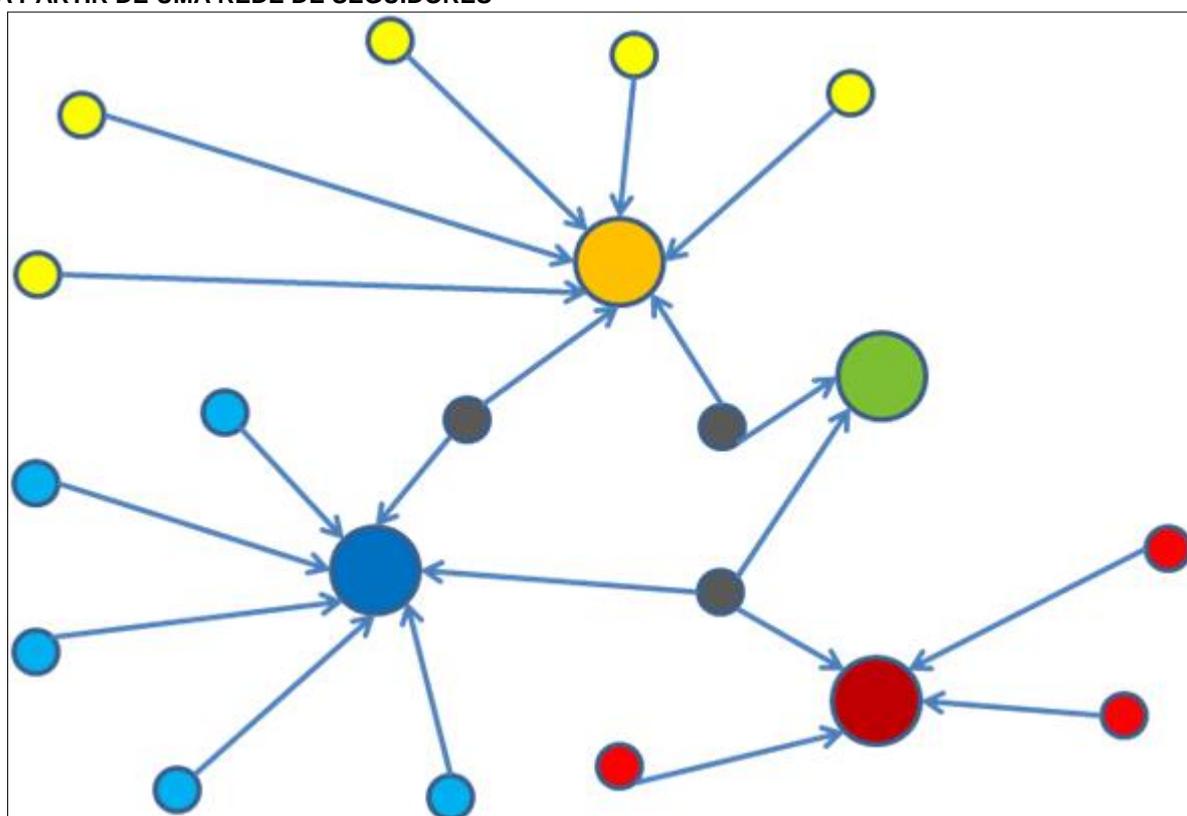
pirâmide de lealdade de Aaker, de modo que um consumidor realmente comprometido com a marca dificilmente fará a escolha por uma marca concorrente.

Aplicando esse conceito às redes sociais, a lealdade a uma marca poderia ser representada pela ação de seguir o perfil da marca a qual o consumidor é leal. Mais do que isso, significa não seguir o perfil daquelas marcas consideradas concorrentes.

Casos extremos de lealdade que se tem observado ultimamente nas redes sociais são em relação a times de futebol ou a partidos políticos. Um usuário realmente leal a um time A e a um partido X seguirá os perfis de A e de X, mas nunca de seus concorrentes diretos B e Y.

A Figura 37 mostra mais uma vez a rede de seguidores de quatro marcas, representadas pelos vértices maiores nas cores amarelo, verde, azul e vermelho. Os vértices menores representam os seguidores de cada uma dessas marcas. Quando esses vértices menores possuem a mesma cor do vértice representativo da marca, significa que há uma relação de exclusividade entre o seguidor e a marca.

**FIGURA 37 – PROPOSTA DO MODELO PREMILINAR PARA ESTABELECIMENTO DA LEALDADE À MARCA A PARTIR DE UMA REDE DE SEGUIDORES**



FONTE: O AUTOR (2015)

Assim, a partir da Figura 37, percebe-se que apesar de as marcas representadas pelas cores amarelo e azul possuírem sete seguidores cada, apenas cinco são seguidores exclusivos, já que dois deles seguem mais de uma marca. A marca verde, por sua vez, não possui nenhum seguidor exclusivo, já que seus seguidores compartilham essa relação com as marcas representadas pelas cores amarela, azul e vermelha. Portanto, as marcas representadas pelos vértices amarelo e azul são aquelas que possuem mais consumidores leais, enquanto a marca representada pela cor verde é a que possui uma menor quantidade de consumidores leais.

Logo, a medida de lealdade a uma determinada marca nas redes sociais poderá ser dada pela seguinte equação:

$$l(x) = seguidores(x) - \sum_{n=1}^N seguidores(x \cap c_n) \quad (20)$$

Onde:

$l(x)$ : lealdade a marca  $x$

$seguidores(x)$ : seguidores da marca  $x$

$N$ : número de concorrentes diretos de  $x$

$c$ : marca concorrente de  $x$

$seguidores(x \cap c_n)$ : seguidores simultâneos de  $x$  e  $c_n$

De forma simplificada, a lealdade absoluta poderia também ser expressa pelo número de seguidores da marca  $x$  cujo grau de saída, em uma rede de seguidores, é igual a 1.

A lealdade normalizada, por sua vez, pode ser expressa por:

$$l'(x) = \frac{l(x)}{seguidores(x)} \quad (21)$$

Onde:

$l'(x)$ : lealdade normalizada à marca  $X$

Ainda com base na Figura 37, percebe-se que apesar de os seguidores das marcas representadas pelos vértices amarelo e azul ser, em números absolutos, mais leais às suas marcas, quando o valor da lealdade é normalizado conforme a proposta apresentada, a marca que possui maior lealdade de seus seguidores é aquela representada pela cor vermelha, uma vez que apesar de ter apenas três seguidores exclusivos, isso representa 75% do número total de seus seguidores.

#### 4.1.3 Qualidade Percebida

Recapitulando mais uma vez o conceito proposto por Aaker, “a qualidade percebida pode ser definida como o conhecimento que o consumidor tem da qualidade geral ou superioridade de um produto ou serviço pretendido, em relação a alternativas” (1998, p. 88). O autor ainda lembra que a qualidade percebida não está necessariamente ligada à qualidade real do produto – que é determinada pela qualidade dos ingredientes, do processo de fabricação, de sua fórmula ou especificações – mas sim à qualidade compreendida pelos próprios consumidores e, portanto, afetada pelos conhecimentos e julgamentos daquilo que cada um deles considera importante.

Esse conceito apresentado por Aaker (1998) é utilizado como base para formulação das questões propostas por Yoo e Donthu (2001) e Pappu, Quester e Cooksey (2005) para medição da qualidade percebida. Esses autores também utilizam a conceituação fornecida por Zeithaml (1988), segundo a qual a qualidade percebida consiste no “julgamento do consumidor sobre a excelência ou superioridade geral de um produto” (ZEITHAML, 1998, p. 3).

Assim, Yoo e Donthu (2001) buscaram medir esse constructo por meio de perguntas como “A qualidade provável da marca X é elevada?” ou “A probabilidade de que X seria funcional é muito alta?” ao passo que Pappu, Quester e Cooksey (2005) perguntaram: “A marca X tem uma boa qualidade?”, “A marca X tem uma qualidade consistente?”, “A marca X é durável? “A marca X é confiável?” e “A marca X tem excelentes características?”.

Transferindo-se esses conceitos para as redes sociais, acredita-se que a qualidade percebida possa ser medida por meio da análise de opinião dos consumidores em relação às marcas. Assim, a qualidade percebida a partir das redes

sociais poderia ser dada pela relação entre os *tweets* positivos e a soma dos *tweets* (positivos e negativos) de determinada marca:

$$qp'(X) = \frac{pos(x)}{pos(x) + neg(x)} \quad (22)$$

Onde:

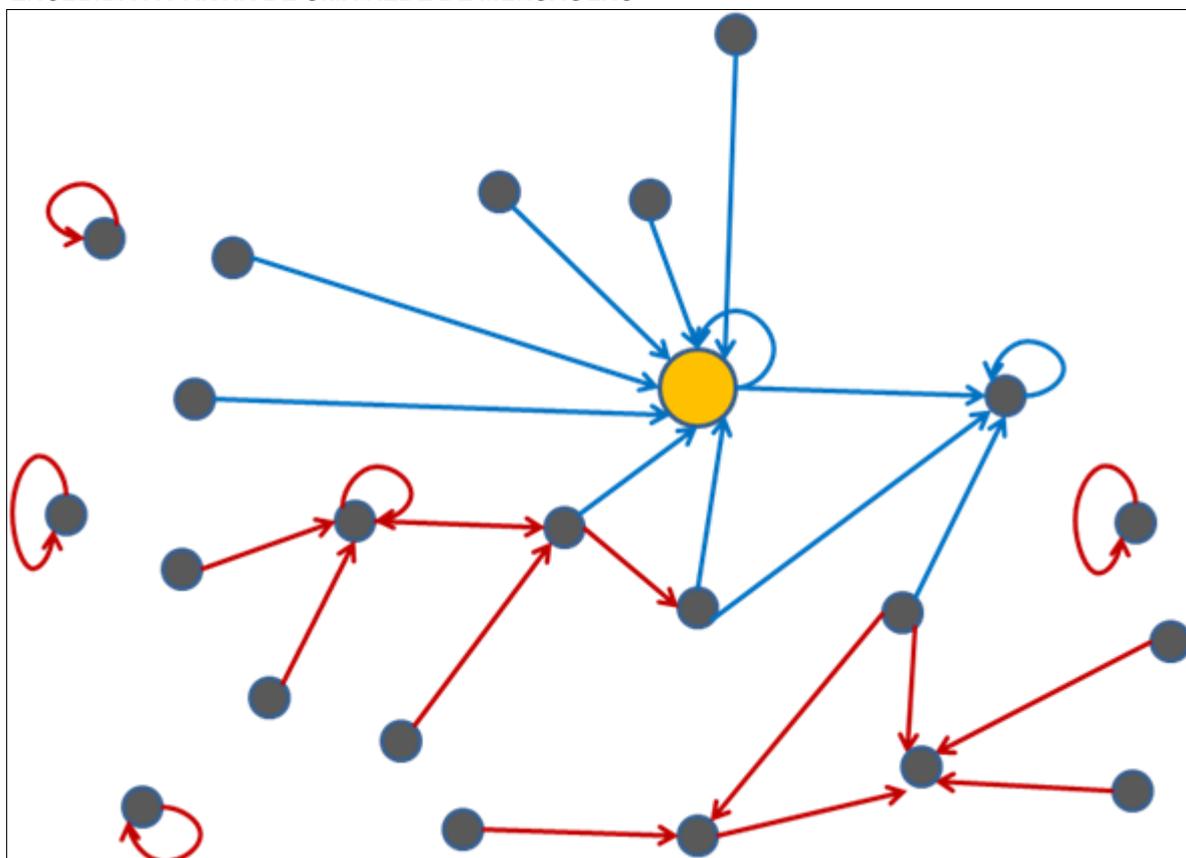
$qp'(x)$ : qualidade percebida normalizada de  $x$

$pos(x)$ : número de *tweets* positivos sobre a marca  $x$

$neg(x)$ : número de *tweets* negativos sobre a marca  $x$

A Figura 38 auxilia na compreensão da proposta. Diferentemente das duas figuras anteriores, não representa uma rede de seguidores, mas sim uma rede de propagação de mensagens na rede.

**FIGURA 38 – PROPOSTA DO MODELO PRELIMINAR PARA ESTABELECIMENTO DA QUALIDADE PERCEBIDA A PARTIR DE UMA REDE DE MENSAGENS**



FONTE: O AUTOR (2015)

Na Figura 38, a publicação de um *tweet* por um perfil da rede é representado por um auto loop (uma aresta que começa e termina no mesmo vértice). Se outro perfil da rede faz *retweet* dessa mensagem, se estabelece então uma nova relação, partindo do perfil que fez o *retweet* para o perfil responsável pela publicação da mensagem original. Da mesma forma, quando um perfil é referenciado através da utilização do símbolo @, cria-se um relacionamento partindo do perfil que realizou a referência para o perfil que foi referenciado.

Percebe-se ainda na Figura 38, que algumas relações aparecem na cor azul e outras na cor vermelha; isso acontece para diferenciar as mensagens com sentimento predominantemente positivo (azul) daquelas com sentimento predominantemente negativo (vermelho). Na mesma figura, o perfil oficial da marca está destacado pela cor amarela.

Assim, podem ser contados na Figura 38, 12 *tweets* positivos em um total de 28 *tweets*. Logo, com base na equação proposta anteriormente, a marca amarela obterá um valor de, aproximadamente, 0,43 para o elemento qualidade percebida, o que indica uma leve predominância de *tweets* negativos em relação aos positivos naquela rede.

Alternativamente, pode-se considerar que a centralidade de cada detentor da opinião influencie o peso dessa opinião na rede como um todo. Assim, uma opinião negativa de um elemento da rede com grande centralidade – determinada, por exemplo, pela medida do PageRank do elemento na rede – pode ter maior impacto na qualidade percebida pelos consumidores do que centenas de opiniões positivas emitidas por membros com menor centralidade. Assim, a qualidade percebida ponderada normalizada pode ser calculada por:

$$qpp'(x) = \frac{\sum_{i=1}^I P(i) * pos(i)}{\sum_{i=1}^I P(i) * pos(i) + \sum_{j=1}^J P(j) * neg(j)} \quad (23)$$

Onde:

$qpp'(X)$ : qualidade percebida ponderada normalizada de x

I: número de elementos que emitiram opinião positiva sobre x

$P(i)$ : PageRank do elemento emissor de opinião positiva

$pos(i)$ : número de mensagens positivas sobre x emitidas pelo nó i

J: número de elementos que omitiram opinião negativa

$P(j)$ : PageRank do elemento emissor de opinião negativa

$neg(j)$ : número de mensagens negativas sobre x emitidas pelo nó j

Assim, a qualidade percebida ponderada de uma marca x é um indicativo do sentimento geral dos membros da rede social em relação à marca, ponderada pela centralidade de cada um desses membros.

#### 4.1.4 Associações à Marca

Conforme define Aaker, “uma associação de marca é algo ‘ligado’ a uma imagem na memória” (AAKER, 1998, p. 114). Portanto, associações são tudo aquilo que o consumidor relaciona com uma marca. Esses relacionamentos, segundo o autor, possuem um grau de força, que é diretamente proporcional à quantidade de experiências ou exposições às comunicações e ações de marketing da empresa. Além disso, quando apoiadas em outras redes de associações anteriormente construídas pela marca, esses relacionamentos se tornam ainda mais fortes na memória do consumidor. Quando há, na mente do consumidor, um conjunto de associações organizadas de forma significativa, está formada, para esse cliente, a imagem da marca (AAKER, 1998).

Nesse sentido, Yoo e Donthu (2001) buscam, em seu modelo operacional, medir às associações às marcas por meio de questões como “Quais as características da marca X vêm rapidamente à sua mente?” ou “Você pode rapidamente se lembrar do logotipo da marca X?”

Portanto, para refletir como é formada a imagem da marca a partir das publicações em uma rede social, essa pesquisa propõe que seja calculada a frequência de cada palavra a partir dos *tweets* publicados sobre determinada marca. Assim, as palavras com maior frequência (após a exclusão das *stopwords*) são aquelas que formam a imagem de uma marca na rede social. Essa imagem pode ser representada graficamente por um artefato de visualização de dados denominado “nuvem de palavras”.

Desta maneira, as associações a marca constituem um elemento qualitativo – e não quantitativo – na avaliação da equidade de uma marca por meio dos dados obtidos das redes sociais.

Para validação do modelo apresentado até aqui, o mesmo foi aplicado aos dados coletados referentes às quatro marcas de cosméticos utilizadas no pré-teste. Os resultados obtidos são apresentados e discutidos na subseção 4.2.

## 4.2 RESULTADOS DA APLICAÇÃO DO MODELO PRELIMINAR

Sobre os dados pré-processados, coletados a partir do Twitter, foi aplicado o modelo proposto para análise de *brand equity*. Os resultados obtidos para os elementos conhecimento da marca, lealdade da marca, qualidade percebida e associações à marca são apontados a seguir.

A Tabela 14 apresenta os resultados obtidos para os componentes conhecimento da marca absoluto ( $c$ ) e conhecimento da marca normalizado ( $c'$ ).

**TABELA 14 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PRELIMINAR PARA O CRITÉRIO CONHECIMENTO DA MARCA REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS**

$X$	<i>followers</i>	$c(x)$	$c'(x)$
Avon	96387	96387	0,376009394
Boticário	98687	98687	0,384981782
Natura	65184	65184	0,254285291
Jequiti	32513	32513	0,126834463

**FONTE: O AUTOR (2015)**

O conhecimento da marca absoluto reflete diretamente o número de seguidores da marca na rede social Twitter, uma vez que é calculado com base no grau de entrada de cada nó da marca na rede de seguidores.

Assim, das quatro marcas analisadas, a mais conhecida é a marca “O Boticário”, representada pelo perfil @oboticario, obtendo o valor 98.687 no critério avaliado, seguida de perto pela marca “Avon”, por meio do perfil @avonbr, que obteve o valor absoluto de 96.387.

Os valores normalizados de conhecimento da marca mostram que a marca “O Boticário” é conhecida por 38,5% dos usuários que compõem a rede analisada, contra 37,6% obtidos pela marca “Avon”. A marca “Jequiti” é acompanhada por apenas 12,7% da rede formada pelos seguidores das marcas estudadas.

Portanto, esses números sugerem que as marcas “O Boticário” e “Avon” são as mais reconhecidas ou lembradas pelos usuários da rede social Twitter dentre as

quatro marcas analisadas. Isso significa que quando um usuário que se interessa por produtos cosméticos e passa a fazer parte dessa rede social, essas são as marcas com maior probabilidade de vir à mente desses consumidores e, conseqüentemente, de serem “seguidas” por eles.

O segundo elemento avaliado foi a lealdade à marca. Os resultados obtidos são apresentados na Tabela 15, em seus valores absolutos e normalizados.

**TABELA 15 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PRELIMINAR PARA O CRITÉRIO LEALDADE À MARCA REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS**

x	l(x)	l'(x)
Avon	51477	0,800097524
Boticário	55196	0,811879984
Natura	16304	0,673570201
Jequiti	15003	0,791591056

**FONTE: O AUTOR (2015)**

A lealdade à marca é representada pelo número de seguidores exclusivos da marca analisada, ou seja, aqueles que seguem a marca, mas não seguem suas concorrentes. Assim, as marcas cujos consumidores são mais leais, tanto quando analisadas em termos absolutos quanto normalizados, são “O Boticário” e “Avon”.

Nesse contexto, as duas marcas com maior conhecimento também foram as duas marcas com maior lealdade. Entretanto, interessante observar a inversão que houve entre as marcas “Natura” e “Jequiti”. Apesar de a marca “Natura” ter obtido melhores índices para o quesito conhecimento da marca e também para lealdade de marca absoluta, quando a lealdade é analisada em termos relativos, “Natura” é superada pela marca “Jequiti”.

Ou seja, apesar de a marca “Natura” ter maior número absoluto de seguidores exclusivos do que a marca “Jequiti”, quando esses valores são normalizados – considerando-se a proporção em relação ao número total de seguidores de cada marca – “Jequiti” obtém um melhor índice de lealdade – 0,79 contra 0,67 da “Natura”.

De um modo geral, os valores normalizados obtidos por meio da análise do critério de lealdade à marca sugerem que os consumidores da marca “O Boticário” são os mais ligados à marca, apresentando menor probabilidade de trocá-la por outra marca, dado que apenas 18,9% deles também seguem pelo menos uma dentre as outras três marcas presentes no estudo. Por outro lado, os seguidores da marca

“Natura”, são os menos “leais”, já que 32,6% também seguem ao menos uma das outras marcas analisadas.

A qualidade percebida, outro critério que compõe a equidade da marca, também foi analisada a partir dos dados extraídos da rede social Twitter. Os valores obtidos para qualidade percebida normalizada ( $qp'$ ) e qualidade percebida ponderada normalizada ( $qpp'$ ) são apresentados na Tabela 16.

**TABELA 16 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PRELIMINAR PARA O CRITÉRIO QUALIDADE PERCEBIDA REFERENTES A CADA UMA DAS QUATRO MARCAS DE COSMÉTICOS ANALISADAS**

x	$qp'(x)$	$qpp'(x)$
Avon	0,424392545	0,411336549
Boticário	0,464276854	0,614634183
Natura	0,830502424	0,854172635
Jequiti	0,142859581	0,549343079

FONTE: O AUTOR (2015)

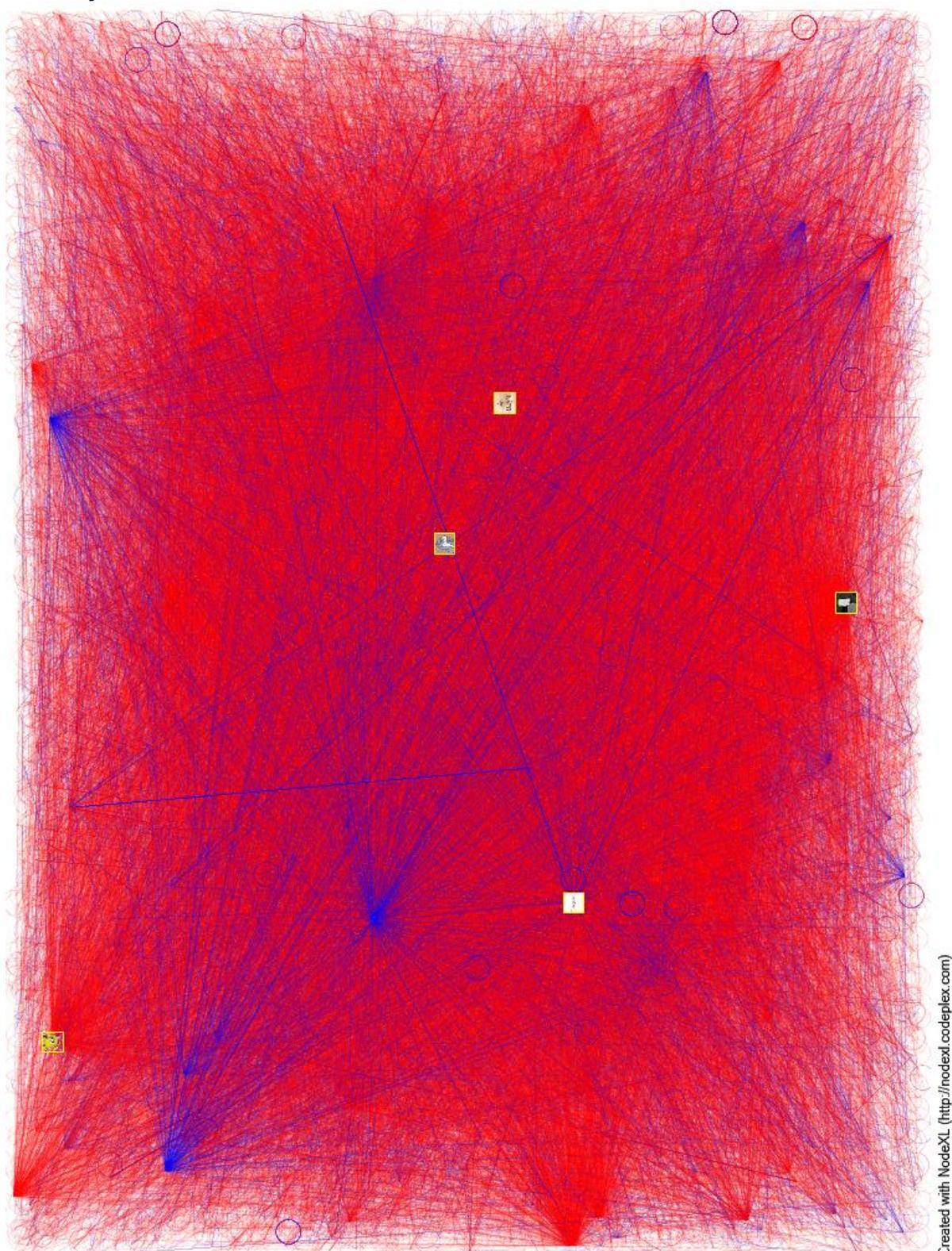
A qualidade percebida reflete o sentimento geral dos membros da rede social Twitter em relação às marcas analisadas, independentemente se esses membros são seguidores da marca ou não. Assim, dentre as marcas analisadas, aquela que obteve maior valor para o critério qualidade percebida foi a “Natura”, atingindo o índice de 0,83. Por outro lado, a menor qualidade percebida apontada para o indicador analisado foi da marca “Jequiti”, com índice de apenas 0,14.

Contudo, quando o valor da qualidade percebida é ponderado pelo prestígio dos emissores dos *tweets* analisados, a marca “Jequiti” supera a marca “Avon”. Isso significa que apesar da maioria dos *tweets* relacionados à marca “Jequiti” ser negativo, aqueles emitidos pelos elementos de maior prestígio na rede são positivos, o que justifica o contraste entre a qualidade percebida e a qualidade percebida ponderada para essa marca.

A Figura 39 mostra a interação entre os usuários do Twitter em relação à marca “Jequiti” e auxilia a compreender esse contraste entre a qualidade percebida e qualidade percebida ponderada para essa marca; os vértices da rede representam os usuários – os cinco com maior PageRank são denotados por sua imagem no perfil da rede e os demais por pontos – enquanto que as arestas representam as mensagens – sendo as azuis representativas de mensagens com sentimento classificado como

predominantemente positivo e as em vermelho representativas de sentimentos classificados predominantemente como negativos.

**FIGURA 39 – REDE DE TWEETS RELACIONADOS À MARCA JEQUITI CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS**



**FONTE: O AUTOR (2015)**

Percebe-se que apesar da maior parte das mensagens carregarem consigo um sentimento predominantemente negativo, aquelas relacionadas a elementos centrais e com maior prestígio na rede, sobretudo emitidas pelo perfil oficial da marca, @jequiti, são predominantemente positivas.

Dos 490 *tweets* escritos pelo perfil @jequiti no período analisado, 332 (ou 67,8%) são positivos e apenas 168 (ou 34,3%) são negativos. Assim, o prestígio desse perfil na rede (PageRank = 748) acaba auxiliando a elevar a qualidade percebida, quando calculada em sua forma ponderada. O Quadro 6 apresenta alguns exemplos de mensagens emitidas pelo perfil, com os respectivos sentimentos predominantes e a confiança obtida no processo de classificação.

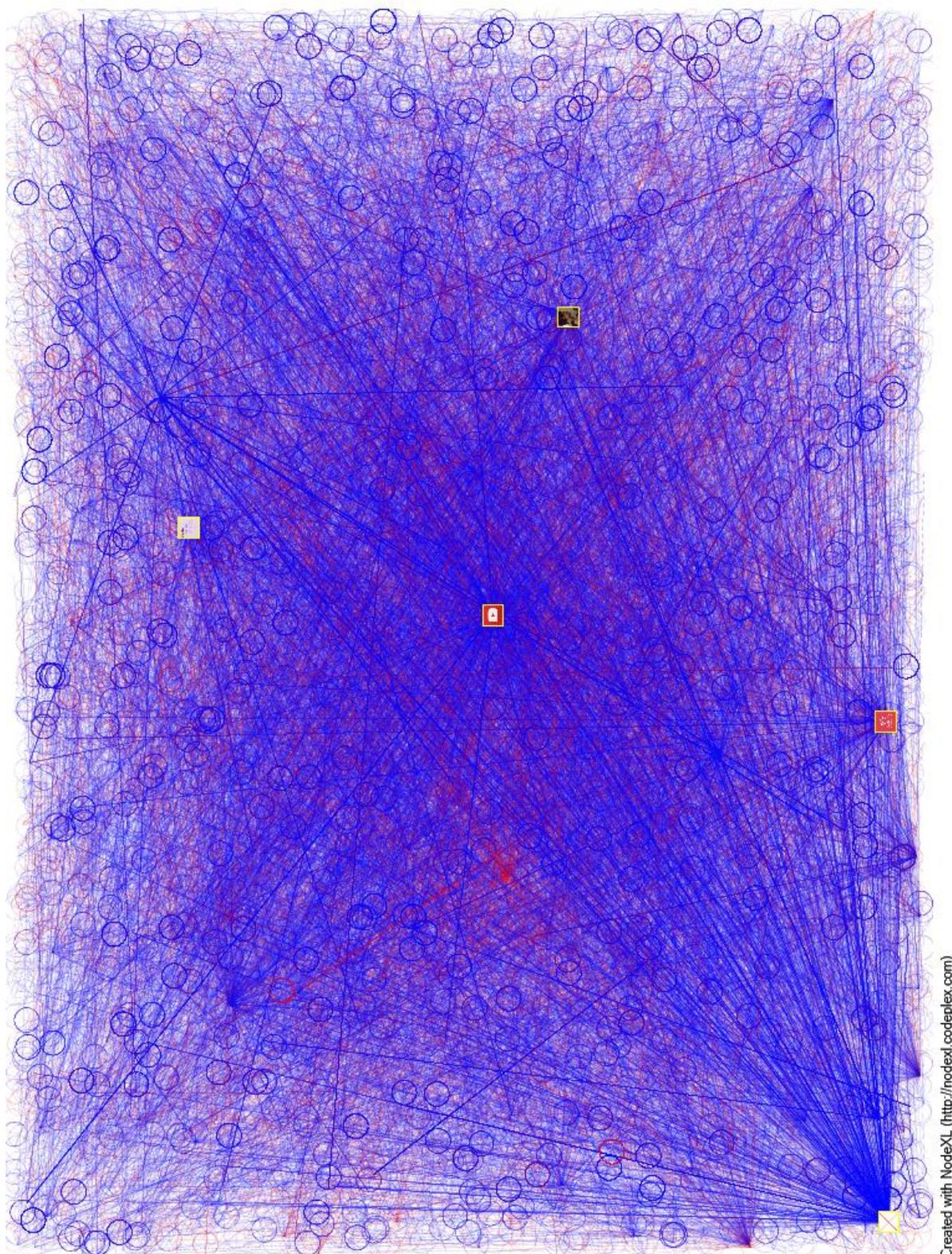
**QUADRO 6 – EXEMPLOS DE TWEETS DO PERFIL OFICIAL DA MARCA JEQUITI**

Origem	Destino	Tipo	Tweet	Sentimento	Confiança
jequiti	jequiti	Tweet	Confira o nosso #MapaOlfativo e conheça os detalhes dos perfumes #Jequiti: <a href="http://t.co/e9SWOCi2Ui">http://t.co/e9SWOCi2Ui</a>	NEGATIVO	0,583157589
jequiti	jequiti	Tweet	Muita sofisticação com #IsabellaFiorentino para #Jequiti. #PerfumesDasEstrelas. <a href="https://t.co/0K7Sw4pETw">https://t.co/0K7Sw4pETw</a>	NEGATIVO	0,758783156
jequiti	jequiti	Tweet	#BlogJequiti: conheça a tendência capilar das próximas temporadas e saiba sobre o #DiaInternacionalDaMulher #Jequiti. <a href="http://t.co/fOI2KnUI20">http://t.co/fOI2KnUI20</a>	POSITIVO	0,627717306
jequiti	sbtonline	Mentions	Hoje tem #RodaARoda #Jequiti! Fique ligado no @SBTonline às 21h45!	NEGATIVO	0,587563419
jequiti	jequiti	Tweet	Com #Jequiti #Aviva suas unhas ficam sempre bonitas e saudáveis! <a href="http://t.co/H5OXTVa6xu">http://t.co/H5OXTVa6xu</a>	POSITIVO	0,517769776
jequiti	jequiti	Tweet	#ConsultorJequiti: acesse o #JequitiComVocê para fazer seus pedidos, ocorrências e acompanhar seus pedidos. <a href="http://t.co/FPUIjoCLt0">http://t.co/FPUIjoCLt0</a>	POSITIVO	0,784689545
jequiti	jequiti	Tweet	No #BlogJequiti falamos sobre o look incrível que a Jennifer Lopez usou no último red carpet: <a href="http://t.co/tsjjJsPNIL">http://t.co/tsjjJsPNIL</a>	POSITIVO	0,754729635
jequiti	jequiti	Tweet	#DicaJequiti: antes do secador e da chapinha, use sempre um produto com proteção térmica.	POSITIVO	0,835754133
jequiti	jequiti	Tweet	A linha #CaminhoDasÁguas ganhou embalagens novas no ciclo 04/2015. Saiba mais no #BlogJequiti: <a href="http://t.co/8tPzNrGyVD">http://t.co/8tPzNrGyVD</a>	POSITIVO	0,780328584

FONTE: O AUTOR (2015)

Percebe-se que apesar de o classificador não supervisionado utilizado atribuir o sentimento predominante corretamente na maior parte das vezes, há erros em alguns casos, como nas duas classificações de sentimento negativo mostradas para os tweets “Confira o nosso #MapaOlfativo e conheça os detalhes dos perfumes #Jequiti: <http://t.co/e9SWOCi2Ui>” e “Hoje tem #RodaARoda #Jequiti! Fique ligado no @SBTonline às 21h45!”. Em ambas, o índice de confiança não chegou a 60%.

FIGURA 40 – REDE DE TWEETS RELACIONADOS À MARCA NATURA CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

A marca “Natura” atingiu um índice de qualidade percebida de 0,83; sua rede é representada na Figura 40. Ao contrário da marca apresentada anteriormente, esse

índice não apresenta variação relevante quando é ponderado pelo prestígio dos emissores das mensagens, chegando a 0,85.

Dos perfis com maior prestígio na rede, dois deles aparentemente são de revendedores da marca, o @ksdescontosecia e @naturaessencial. Há também um número relevante de mensagens emitidas pelo perfil @naturanet, sendo a maioria desses *tweets* do tipo “Replies To”, mostrando a constante interação da empresa com os usuários da rede social por meio do perfil oficial da marca. Nota-se ainda o prestígio do perfil “Youtube” nessa rede; entretanto, é uma participação passiva, já que todos os *tweets* relacionados ao perfil são do tipo “Mentions”, ou seja, o perfil não é o emissor da mensagem, mas é citado por outros usuários em suas publicações. Exemplos de *tweets* desses perfis com maior prestígio na rede são apresentados no Quadro 7.

**QUADRO 7 – EXEMPLOS DE TWEETS DOS PERFIS COM MAIOR PRESTÍGIO NA REDE DA MARCA NATURA**

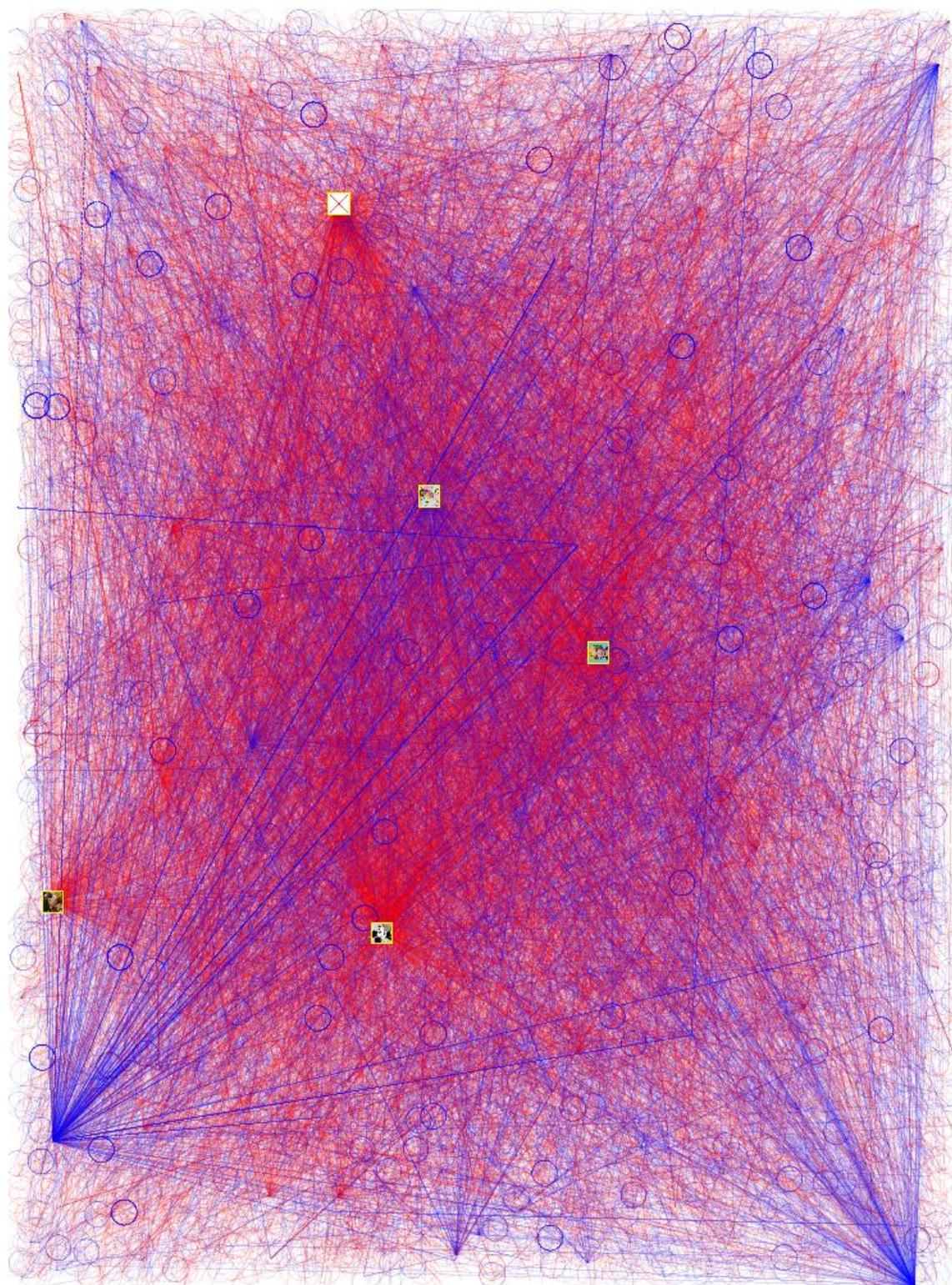
Origem	Destino	Tipo	Tweet	Sentimento	Confiança
Ksdescontosecia	ksdescontosecia	Tweet	#diadamulher #descontosecia #natura mais 15 % de desconto com o cupom REDENATURA oferta ▶ <a href="http://t.co/CAoldJ1Pa1">http://t.co/CAoldJ1Pa1</a> <a href="http://t.co/lmTvn7Jvj7">http://t.co/lmTvn7Jvj7</a>	POSITIVO	0,691499237
Ksdescontosecia	ksdescontosecia	Tweet	Só hoje mais 15% de desconto em todo o site usando o o cupom REDENATURA <a href="http://t.co/wWcwMh52BK">http://t.co/wWcwMh52BK</a> <a href="http://t.co/AlyO4MG25o">http://t.co/AlyO4MG25o</a>	POSITIVO	0,599251028
Naturaessencial	naturaessencial	Tweet	• Promoção Óleo Desodorante Corporal Avelã Sève De R\$ 64,90 por R\$ 34,40 <a href="http://t.co/AzBUbDod0l">http://t.co/AzBUbDod0l</a> Entrega em casa. <a href="http://t.co/zokgDVSEIC">http://t.co/zokgDVSEIC</a>	POSITIVO	0,817265227

Naturaessencial	anaventuran	Mentions	Olá @anaventuran Sou Consultora Digital Natura No meu espaço de vendas você compra com segurança, sem sair de casa <a href="http://t.co/XSqUG2PZjG">http://t.co/XSqUG2PZjG</a>	POSITIVO	0,504868431
Naturanet	distrito_07	Replies to	@Distrito_07 para peles oleosas, a gente indica a base Mousse efeito pó da linha Una. você já conhece? (cont.) <a href="http://t.co/xvBskTKQgg">http://t.co/xvBskTKQgg</a>	POSITIVO	0,550072977
Naturanet	fabianefarias6	Replies to	@FabianeFarias6, o água de banho rara priprioca ainda está disponível! Olha ele aqui: <a href="http://t.co/NihOCNj1a6">http://t.co/NihOCNj1a6</a> #ChameQueVem 😊	POSITIVO	0,571307814
Naturanet	luris157	Replies to	@luris157, o perfume mais concentrado da Natura é o Perfume do Brasil de Priprioca Ekos, seguido do Natura Una. 😊#ChameQueVem	NEGATIVO	0,52830464
Naturanet	diretodoalem	Replies to	@diretodoalem, para tirar uma maquiagem mais pesada dos olhos, o demaquilante bifásico é o ideal! 😊 <a href="http://t.co/KmoDbdjiTl">http://t.co/KmoDbdjiTl</a> #ChameQueVem	NEGATIVO	0,518298242
baby119glamour	youtube	Mentions	Gostei de um vídeo @YouTube <a href="http://t.co/7NopWqQNg6">http://t.co/7NopWqQNg6</a> caixa natura ciclo 03 2° pedido	POSITIVO	0,752297977

**FONTE: O AUTOR (2015)**

A marca “Avon”, por outro lado, apresenta predominância de mensagens negativas tanto quando a qualidade percebida é avaliada de forma não ponderada quanto ponderadamente, obtendo os índices de 0,42 e 0,41, respectivamente. A rede apresentada na Figura 41 reflete esse cenário.

FIGURA 41 – REDE DE TWEETS RELACIONADOS À MARCA AVON CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

Como se pode perceber pela Figura 41 e também pelo Quadro 8, nos perfis com maior prestígio prevalecem mensagens com sentimento predominantemente negativo.

**QUADRO 8 – EXEMPLOS DE TWEETS DOS PERFIS COM MAIOR PRESTÍGIO NA REDE DA MARCA AVON**

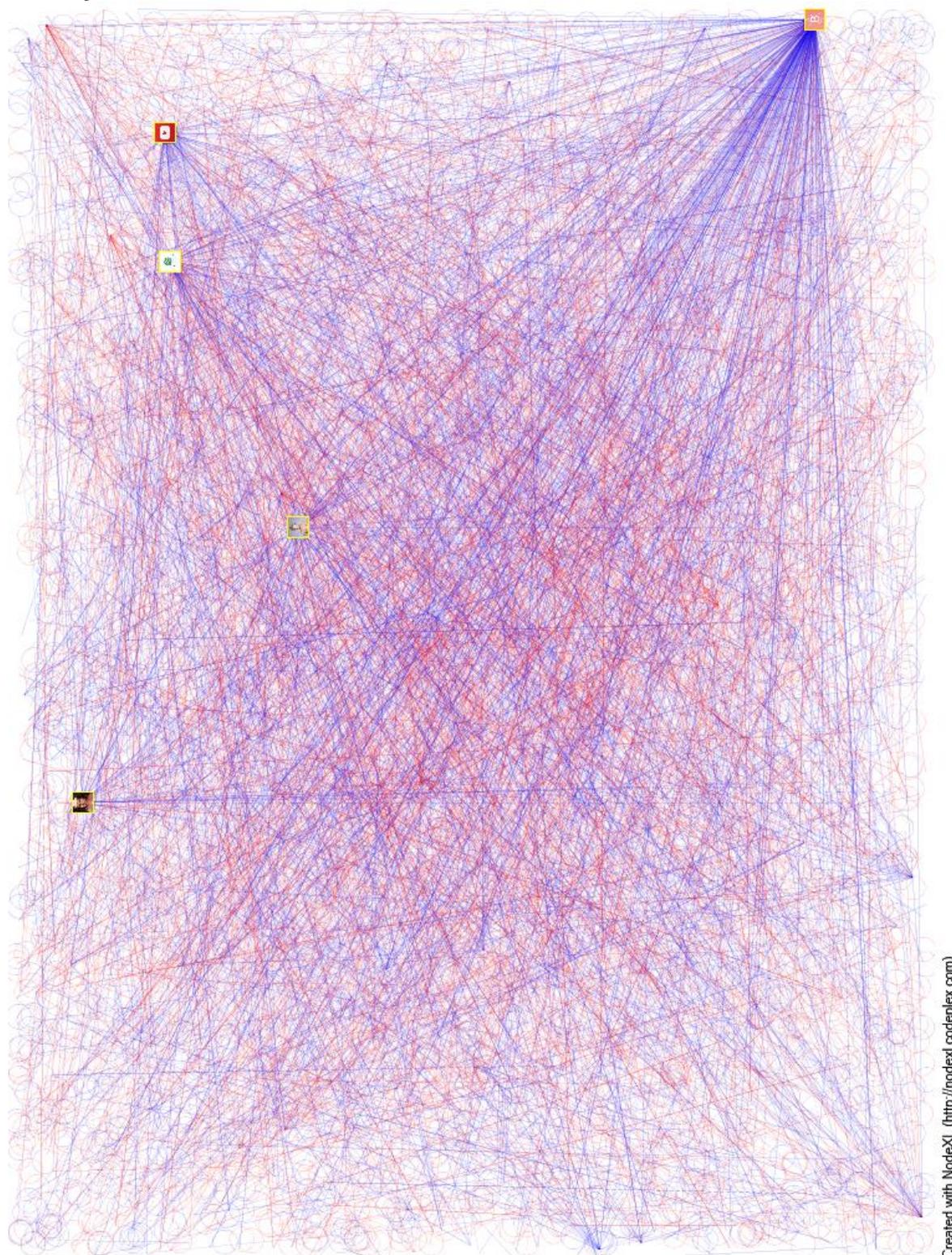
Origem	Destino	Tipo	Tweet	Sentimento	Confiança
instagranzin	instagranzin	Tweet	abrir o caderno de matematica e ficar naquela duvida imensa: ser revendedor jequiti ou avon	NEGATIVO	0,68580127
luscaspfvr	pqpvoce	Replies to	@pqpvoce desculpa nunca usei esse eu esfrego meu pulso nas revista da avon	NEGATIVO	0,543775317
luscaspfvr	luscaspfvr	Tweet	oi mae comprei um perfume otimo pra sinhora da avon aqui oh ta aqui na revistinha so esfregar no pulso funciona mesmo te amo	NEGATIVO	0,818689931
zoando	zoando	Tweet	abrir o caderno de matematica e ficar naquela duvida imensa: ser revendedor jequiti ou avon	NEGATIVO	0,68580127
Frasesdebebeada	frasesdebebeada	Tweet	a vodka deixa as pessoas tão mais bonitas que daqui à pouco será vendida pela avon	NEGATIVO	0,638913444
harry_momento	harry_momento	Tweet	Valendo um kit da Avon ! #KCA #VoteUKDirectioners <a href="http://t.co/vnSY4aFuqt">http://t.co/vnSY4aFuqt</a>	POSITIVO	0,841594042

**FONTE: O AUTOR (2015)**

Mais uma vez, conforme se pode observar no Quadro 11, o classificador não-supervisionado de sentimento se comporta de forma satisfatória, acertando o sentimento predominante nos *tweets*.

Por fim, apresenta-se a rede obtida referente a marca “O Boticário” na Figura 42; percebe-se que é uma rede menos densa se comparada as outras três, o que reflete um menor número de *tweets* referentes a marca em comparação as demais presentes no estudo.

**FIGURA 42 – REDE DE TWEETS RELACIONADOS À MARCA O BOTICÁRIO CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS**



**FONTE: O AUTOR (2015)**

A marca “O Boticário”, assim como “Jequiti”, apresenta um crescimento significativo de qualidade percebida quando o índice é ponderado pelo prestígio do

agente emissor; como se observa na Figura 42, apesar do número de *tweets* predominantemente negativos ser um pouco maior do que os predominantemente positivos, os elementos com maior prestígio na rede emitem mais mensagens positivas do que negativas. O Quadro 9 mostra os *tweets* que envolvem os perfis com maior prestígio na rede.

**QUADRO 9 – EXEMPLOS DE TWEETS DOS PERFIS COM MAIOR PRESTÍGIO NA REDE DA MARCA O BOTICÁRIO**

Origem	Destino	Tipo	Tweet	Sentimento	Confiança
oboticario	Oboticario	Tweet	Nos dias 26 e 27/02, diversos produtos com 15% de desconto na nossa loja online. Vem conferir! <a href="http://t.co/vyqfZ642MU">http://t.co/vyqfZ642MU</a> <a href="http://t.co/vcETVwMAX0">http://t.co/vcETVwMAX0</a>	POSITIVO	0,729632245
oboticario	barbosagisele	Replies to	@BarbosaGisele Agora você pode trocar os pontos na Loja Online também. Confira o regulamento <a href="https://t.co/zN04ixl35z">https://t.co/zN04ixl35z</a> e aproveite! ;)	POSITIVO	0,752409351
oboticario	suelenjohann	Mentions	Já viu os vídeos que preparamos para você com a nossa maquiadora @SuelenJohann? Confira!!	POSITIVO	0,707954266
oboticario	siimone_1971	Replies to	@siimone_1971, Oi! Para atendê-la com assertividade, por favor, nos passe detalhes pelo Chat <a href="http://t.co/aoyPfhw6zY">http://t.co/aoyPfhw6zY</a> ou atendimento privado.	NEGATIVO	0,507361851
nadaperfeita_	Youtube	Mentions	Blush liquido O Boticario-Resenha: <a href="http://t.co/X73OjORFHA">http://t.co/X73OjORFHA</a> via @YouTube	POSITIVO	0,632548159
oboticario	suelenjohann	Mentions	Já viu os vídeos que preparamos para você com a nossa maquiadora @SuelenJohann? Confira!! <a href="http://t.co/g8ewnPxOBi">http://t.co/g8ewnPxOBi</a> <a href="http://t.co/QafHAztggB">http://t.co/QafHAztggB</a>	POSITIVO	0,707954266
_crljuu	suelenjohann	Mentions	RT @oBoticario: Já viu os vídeos que preparamos para você com a nossa maquiadora @SuelenJohann? Confira!! <a href="http://t.co/g8ewnPxOBi">http://t.co/g8ewnPxOBi</a> <a href="http://t.c...">http://t.c...</a>	POSITIVO	0,692773852
eukarlinha	Eukarlinha	Tweet	A moça na porta da Boticário sou eu <a href="http://t.co/CxVy8nV5Zi">http://t.co/CxVy8nV5Zi</a>	POSITIVO	0,605450059

FONTE: O AUTOR (2015)

Assim como a “Natura”, a marca “O Boticário” faz uso de seu perfil oficial no Twitter para interagir com os usuários da rede, como no *tweet* “@BarbosaGisele Agora

you can exchange the points in the Online Store as well. Confirm the regulation <https://t.co/zN04ixl35z> and enjoy! ;)” shown in Table 12. There is also *tweets* mentioning the profile @youtube when making a reference to some video shared on the site. The profile of @suelenjohan is also mentioned on the network, advanced mainly by the citations made by the profile @oboticario.

As associações às marcas, que buscaram identificar as principais palavras que o consumidor relaciona com as marcas estudadas, foram o último componente de *brand equity* analisado. Com base nos *tweets* coletados de cada uma das marcas, foram levantadas as palavras com maior frequência, cujo resultado é apresentado em dois formatos: por meio de ranking, com as dez palavras mais frequentes, e visualmente, por meio de uma nuvem de palavras, na qual o tamanho das palavras também é proporcional a sua frequência.

Para cada marca presente na pesquisa, foram analisados três cenários: um levando em consideração todos os *tweets* coletados da marca, outro considerando apenas os *tweets* que foram classificados como positivos e outro considerando apenas os negativos. Os resultados são apresentados a seguir.

O - PER0 mostra as palavras com maior frequência nos *tweets* da marca Avon:

**QUADRO 10 – PALAVRAS MAIS FREQUENTES NOS TWEETS RELACIONADOS À MARCA AVON**

Geral		Positivos		Negativos	
Avon	25142	Avon	10525	Avon	14617
Jequiti	3591	Revista	935	Jequiti	3394
Ser	3341	Eu	824	Ser	3042
Eu	2744	Vender	688	Ficar	2546
Ficar	2613	Só	676	Revendedor	2377
Abrir	2527	Produtos	616	Abri	2309
Revendedor	2443	Natura	608	Caderno	2307
Caderno	2308	Youtube	590	Duvida	2306
Duvida	2307	Meu	560	Imensa	2301
Imensa	2301	Video	513	Naquela	2300

**FONTE: O AUTOR (2015)**

Como esperado, a palavra “Avon” é a mais frequente, já que os *tweets* analisados são sobre essa marca; entretanto, outras marcas, como “Jequiti” e “Natura” também aparecem nos rankings. A Figura 43 mostra as palavras mais associadas à marca “Avon”.





Por outro lado, quando as palavras “Avon” e “Jequiti” aparecem associadas aos *tweets*, geralmente revelam um sentimento negativo, na medida em que o emissor da mensagem associa a tarefa de vender produtos dessas marcas como sendo a última alternativa, como em “abro o livro pra estudar já me decidindo se vou vender Avon ou Jequiti” ou em “Aah sim, quero medicina, caso não dê, quero odontologia, caso não dê, vou vender jequiti, avon e natura kkkk” ou ainda em “abrir o caderno de matematica e ficar naquela duvida imensa:ser revendedor jequiti ou Avon”.

Três das palavras utilizadas na frase anterior – “duvida” e “imensa” e “revendedor” – também aparecem no ranking, sendo associadas mais a *tweets* negativos do que positivos, aparecendo em destaque na Figura 45.

FIGURA 45 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE AVON CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

Observa-se a partir da Figura 45 que palavras como “matemática”, “pulso” e “esfregar” também são bastante associadas com *tweets* negativos da marca Avon,

como exemplificam os *tweets* “Abri o caderno de matemática e fica aquela duvida imensa ser revendedor da Jequiti ou Avon?”, “vcs tem cara q passa o pulso nas revista Avon” e “Seu recalque bate no meu perfume importado e volta como revista da avon, pra vc esfregar no pulso”.

O Quadro 11, por sua vez, apresenta as palavras mais frequentemente relacionadas aos *tweets* da marca “O Boticário”.

**QUADRO 11 – PALAVRAS MAIS FREQUENTES NOS TWEETS RELACIONADOS À MARCA O BOTICÁRIO**

Geral		Positivos		Negativos	
Boticario	7575	Boticario	3273	Boticario	4302
Perfume	1274	Oboticario	369	Perfume	1110
Meu	1091	Eu	307	Meu	390
Me	503	Voce	306	Me	356
Você	490	Dia	273	Minha	326
Oboticario	431	Já	230	Comprar	295
Tem	420	Make	189	Vou	276
Dia	417	Perfume	164	Tem	267
Minha	415	Confira	158	Mais	247
Mais	412	Tem	150	Avon	241

**FONTE: O AUTOR (2015)**

Destacam-se as associações com as palavras “make”, predominante em *tweets* positivos, como em “Linda essa make para esse final de semana vem ver no blog (usei a sombra green -o boticario)...”, além das palavras “perfume” e a concorrente “Avon”, predominantes em *tweets* negativos sobre a marca “O Boticário, como em “Ninguém me tira da cabeça que o perfume glamour da boticário tem cheiro de hipoglós” e “comprei toda a linha matte da boticário e alguns da avon e fui roubada no mesmo dia, não tá fácil, amigas”. As principais associações com a marca “O Boticário” são apresentadas na Figura 46.







QUADRO 12 – PALAVRAS MAIS FREQUENTES NOS TWEETS RELACIONADOS À MARCA JEQUITI

Geral		Positivos		Negativos	
Jequití	60439	Jequití	7933	Jequití	52506
Casa	9744	Roda	4775	Casa	9719
Roda	9720	Produtos	2247	Prova	9673
Prova	9675	Valendo	1014	Aula	9618
Aula	9621	Existe	981	Dever	9556
Dever	9557	Mulher	958	Perfume	7436
Perfume	7610	Showdomilhao	836	Luansantana	5213
Produtos	6666	Cunha	798	Roda	4945
Luansantana	5219	Cota	796	Se	4868
Perfumes	4827	Aprova	795	Perfumes	4810

FONTE: O AUTOR (2015)

Observam-se ocorrências importantes das palavras “casa”, “roda”, “prova”, “aula” e “luansantana”. Todas aparecem em destaque na Figura 49.

FIGURA 49 – NUVEM DE PALAVRAS ASSOCIADAS À MARCA JEQUITI CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

Apesar da palavra “roda” ser associada, em números absolutos, mais à *tweets* negativos do que positivos, é, proporcionalmente, mais frequente nos positivos, aparecendo, portanto, em destaque na Figura 50. Essa palavra é relacionada ao programa “Roda a Roda Jequiti”, apresentado por Silvo Santos no canal Sbt, como nos *tweets* “*Esse Roda a Roda Jequiti eu acertei de prima*”, “*vdd. E Roda a Roda Jequiti é um dos melhores programas da TV brasileira por razoes de: tem Silvio Santos*” e “*Trívia: com assinatura vitalícia do Netflix, Silvio Santos vai oferecer o login dele ao invés de dinheiro no Roda a Roda Jequiti.*”





outras palavras presentes na nuvem de palavras associadas a *tweets* negativos sobre Jequití.

Por fim, as palavras mais frequentemente associadas aos *tweets* da marca “Natura” também foram analisadas. Os resultados são apresentados no Quadro 13.

**QUADRO 13 - PALAVRAS MAIS FREQUENTES NOS TWEETS RELACIONADOS À MARCA NATURA**

Geral		positivo		negativo	
Natura	36839	Natura	29544	Natura	7295
Dia	6752	Dia	6504	Dei	976
Promoções	4552	Promoções	4533	Eu	805
Ani	4233	Jz	3990	Meu	776
Jz	4012	Presente	3553	Avon	756
Presente	3792	Promoção	3380	Perfume	711
Promoção	3486	Aproveite	3148	Se	669
Aproveite	3389	Desconto	2846	Feliciano	632
Tem	3175	Tem	2734	Mais	461
Desconto	3015	Mães	2694	Jequiti	446

**FONTE: O AUTOR (2015)**

Nota-se pela primeira vez, dentre todas as marcas analisadas, a ocorrência da palavra “Mães” nos *tweets* positivos, sugerindo, em composição com a palavra “Dia”, também bastante presente nesses *tweets*, que a marca “Natura” tenha sido mais associada ao dia das mães do que suas concorrentes, como exemplificam os *tweets* “*Bom dia galera! O dia das mães está chegando e a Natura já pensou em tudo, preparou kits incríveis pra você...*”, “*O mês que se comemora o dia das mães já esta se aproximando e a Natura já lançou suas novas promoções, vamos começar a conferir!*” e “*Amanhã gastarei o vale presente que ganhei da natura por escrever uma dedicatória pra minha mãe para o dia das mães*”. Todas essas palavras podem ser observadas na Figura 52.

FIGURA 52 – NUVEM DE PALAVRAS ASSOCIADAS À MARCA NATURA CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

A Figura 53, por sua vez, apresenta as relações mais comuns encontradas nos tweets com sentimento predominantemente positivo relacionados à marca, merecendo destaque as palavras “promoção” (além de sua flexão no plural), “presente” e “desconto”.





A palavra “boicote”, que pode ser observada tanto na nuvem de palavras positivas quanto na de negativas, também está relacionada aos *tweets* sobre a repercussão da polêmica declaração do deputado Marco Feliciano.

#### 4.3 CONSIDERAÇÕES SOBRE O MODELO PRELIMINAR

Após a aplicação do modelo preliminar sobre os dados coletados do Twitter e a análise dos resultados obtidos, algumas considerações foram feitas para identificar pontos positivos e negativos do modelo e, a partir dessa avaliação, mudanças foram incorporadas ao modelo proposto.

Os resultados obtidos na análise do *brand equity* sob a perspectiva do consumidor a partir das mídias sociais parecem evidenciar que o modelo preliminar apresentado atingiu o objetivo para o qual foi concebido, na medida em que possibilitou a análise dos elementos que compõem a equidade da marca – conhecimento da marca, lealdade à marca, qualidade percebida e associações à marca – inclusive de forma quantitativa e normalizada para três dos quatro fundamentos analisados.

Dessa forma, o modelo preliminar se assemelha aos apresentados por Yoo e Donthu (2001) e Pappu, Quester e Cooksey (2005), na medida em que pôde mensurar como as marcas são percebidas por seus consumidores, operacionalizando, assim, os modelos conceituais propostos por Keller (1993) e Aaker (1998). Entretanto, essa pesquisa se difere dos modelos operacionais apresentados anteriormente em dois pontos importantes: no modo como a coleta dos dados foi realizada e nas técnicas empregadas para análise desses dados.

Os modelos operacionais propostos anteriormente podem ser dispendiosos e inconvenientes, à medida que utilizam questionários para coleta de dados. No modelo aqui apresentado, por outro lado, foram utilizados dados coletados da rede social Twitter, cuja manifestação dos usuários é realizada de forma não incitada. Destarte, garante-se que as opiniões e a ação de seguir os perfis relacionados às marcas sejam realizadas por livre e espontânea vontade do usuário, sem nenhuma influência ou interferência do pesquisador.

Em relação à análise dos dados, ao invés da utilização de complexos métodos estatísticos que podem ser de difícil compreensão por parte dos profissionais da área

de marketing, o modelo preliminar resultante dessa pesquisa faz uso de técnicas de análise de redes sociais e mineração de dados.

Algumas oportunidades de melhoria foram identificadas a partir da análise dos resultados do modelo preliminar. A avaliação das marcas sob a perspectiva do consumidor pode ter sido influenciada por perfis não consumidores, presentes nas redes sociais para representar marcas, vendedores, sites jornalísticos e outras “pessoas jurídicas” que não são, necessariamente, consumidores. Isolar tais elementos da análise, sobretudo em relação à qualidade percebida e às associações à marca, foi um dos desafios do modelo proposto.

Em seu trabalho para verificar como o valor financeiro de uma marca pode ser influenciado pelo conhecimento da marca e imagem da marca, Nam e Kannan (2014) utilizaram rótulos coletados da mídia social *Delicious* para inferir sobre esses elementos. Ao passo que o conhecimento foi medido indiretamente pelo volume de rótulos publicados sobre determinada marca, a imagem da marca é analisada por meio de associações das marcas com esses rótulos, o que é mostrado por meio de uma rede formada pelas marcas e os rótulos a elas associados.

Nesse mesmo trabalho, os autores classificam as associações em sete categorias: atitudes ou avaliações (ex: legal, bem-humorado, inovador, criativo) atributos do produto (ex: acessibilidade, confiabilidade, estabilidade), atributos não relacionados ao produto (ex: preço, serviço), categoria do produto (ex: televisão, brinquedos, hotel), marcas (ex: Apple, Microsoft, iPod, Google), nomes próprios (ex: Curitiba, Nova York, Brasil, Eduardo) e palavras descritivas (ex: artigo, comportamento, negócios, ônibus).

Essa categorização das associações possibilitou um *insight* visando a melhoria do modelo preliminar: a análise morfológica das palavras (classificando-as em substantivo, artigo, adjetivo, numeral, pronome, verbo, advérbio, preposição, conjunção e interjeição) pode ser útil principalmente para melhoria da proposta relacionada a qualidade percebida. Assim, a qualidade percebida tornou-se, no modelo proposto, concernente aos adjetivos relacionados à marca pesquisada e um novo elemento foi criado para contemplar o sentimento geral observado a partir das publicações dos usuários na rede social.

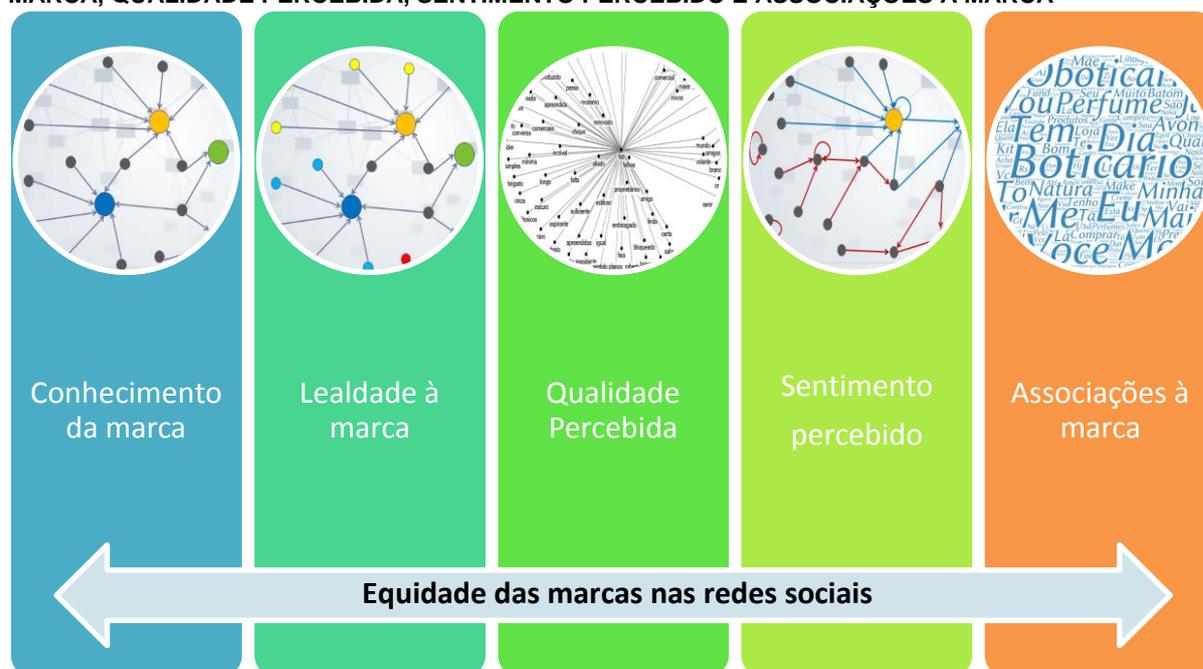
Outro desafio do modelo proposto foi sugerir uma forma de transformar os resultados quantitativos do modelo preliminar referentes aos constructos lealdade à

marca e conhecimento da marca em valores qualitativos, adequando-os às pirâmides da lealdade e do conhecimento apresentadas por Aaker (1998).

#### 4.4 MODELO PROPOSTO

O modelo proposto para análise da equidade das marcas a partir dos dados provenientes da rede social Twitter apresenta algumas modificações em relação ao modelo preliminar. A representação gráfica desse modelo pode ser observada na Figura 55.

**FIGURA 55 – ESQUEMATIZAÇÃO DO MODELO PROPOSTO PARA ANÁLISE DA EQUIDADE DAS MARCAS NAS REDES SOCIAIS, CONTEMPLANDO CINCO DIMENSÕES: CONHECIMENTO DA MARCA, LEALDADE À MARCA, QUALIDADE PERCEBIDA, SENTIMENTO PERCEBIDO E ASSOCIAÇÕES À MARCA**



FONTE: O AUTOR (2015)

Nas subseções a seguir são apresentadas as diferenças de cada um desses elementos em relação ao modelo preliminar.

##### 4.4.1 Conhecimento da marca

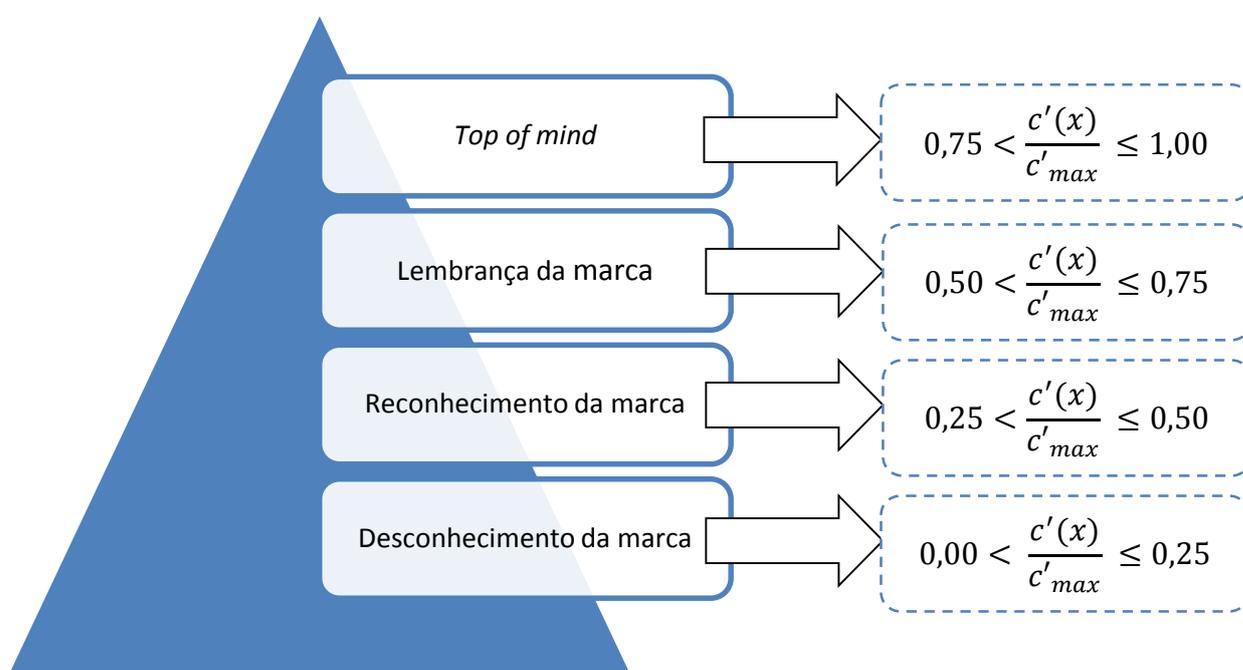
Assim como no modelo preliminar, o conhecimento da marca se baseia na premissa de que há uma relação direta entre o número de seguidores de uma marca e o conhecimento dessa marca na rede social. Portanto, fazendo o uso das

ferramentas de análise de redes sociais, pode-se afirmar que esse conhecimento pode ser enunciado em termos do grau de entrada do vértice que representa uma marca, considerando sua rede de seguidores.

Com base nos resultados de aplicação do modelo preliminar obtidos para o critério conhecimento da marca (Tabela 14), pode-se observar que os valores normalizados para esse critério variam, aproximadamente, de 0,13 a 0,38. Entretanto, considerando um cenário extremo em que nenhum dos membros da rede seguem uma marca pesquisada e, ao mesmo tempo, todos seguem uma marca concorrente, dentro da mesma categoria de produtos, poder-se-ia ter o valor normalizado do conhecimento da marca variando entre 0 e 1.

Assim, para adequar esses valores à pirâmide de conhecimento da marca proposta por Aaker (1998) – Figura 4 – propõe-se uma transformação linear dos valores de conhecimento normalizado da marca para corresponder a cada um dos quatro níveis da pirâmide. A equivalência entre os valores obtidos para o conhecimento da marca normalizado e a pirâmide do conhecimento de Aaker (1998) é mostrada na Figura 56.

**FIGURA 56 – CORRESPONDÊNCIA ENTRE O CONHECIMENTO DA MARCA NORMALIZADO E OS NÍVEIS DE CONHECIMENTO DE AAKER (1998)**



FONTE: O AUTOR (2015)

Retomando os resultados normalizados obtidos para o conhecimento das marcas de cosméticos pesquisadas, as marcas Avon e Boticário, de acordo com a proposta do novo modelo, seriam associadas à categoria *Top of Mind*, enquanto que Natura seria classificada como uma marca lembrada e Jequití apenas como uma marca reconhecida.

#### 4.4.2 Lealdade à marca

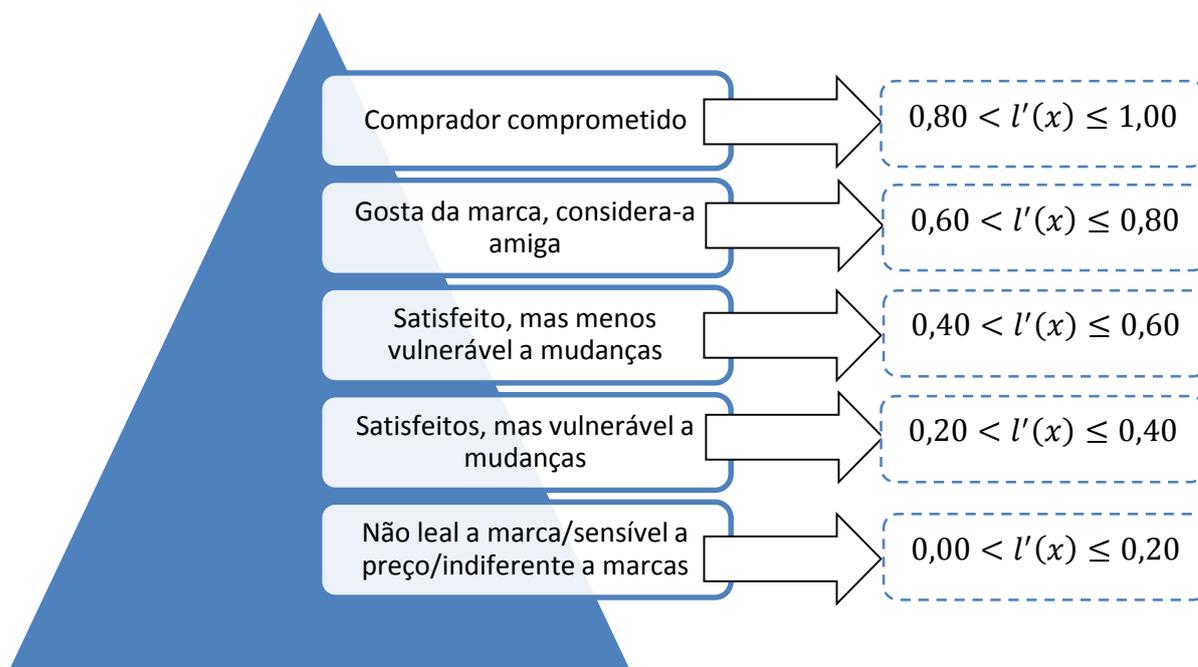
O modelo preliminar propõe que a lealdade à marca, no contexto das redes sociais, pode ser determinada pela exclusividade de seus seguidores, ou seja, quanto maior o número de usuários que segue o perfil de uma marca e não segue seus concorrentes, maior será o nível de exclusividade e, portanto, maior a lealdade do consumidor à marca.

Assim, a lealdade absoluta pode ser expressa pela diferença entre o número de seguidores de uma marca e o somatório do número de seguidores simultâneos dessa marca e seus concorrentes. A lealdade normalizada, por sua vez, pode ser obtida pela razão entre a lealdade absoluta e o número total de seguidores da marca.

Analisando os resultados obtidos pela aplicação desse conceito sobre as marcas pesquisadas de cosméticos (Tabela 15), os valores para a lealdade ficaram entre 0,67 e 0,81. Contudo, considerando uma situação limite, poder-se-ia ter esse valor variando entre 0 e 1, respectivamente, nos seguintes casos: a) todos os seguidores de uma marca seguem também pelo menos um de seus concorrentes; b) todos os seguidores da marca a seguem com exclusividade e, portanto, não segue nenhum de seus concorrentes.

A partir disso, para o modelo proposto, sugere-se uma conversão linear entre os possíveis valores da lealdade normalizada da marca e os cinco níveis de lealdade apresentados por Aaker (1998). Essa correspondência é mostrada na Figura 57.

FIGURA 57 – CORRESPONDÊNCIA ENTRE A LEALDADE NORMALIZADA À MARCA E OS NÍVEIS DE LEALDADE DE AAKER (1998)



FONTE: O AUTOR (2015)

Aplicando a correspondência proposta na Figura 57 aos resultados obtidos para lealdade normalizada referente às marcas de cosméticos, haveria duas marcas – Natura e Jequiti – na categoria “Satisfeito, mas vulnerável a mudanças” e outras duas – Boticário e Avon – na categoria “Gosta da marca, considera-a amiga” mostrando que, pelo menos nas redes sociais, os potenciais consumidores das marcas Boticário e Avon são mais leais a essas marcas do que aqueles que seguem as marcas Natura e Jequiti.

#### 4.4.3 Qualidade percebida

Conforme proposta do modelo preliminar, a qualidade percebida era calculada a partir da análise do sentimento predominante nos *tweets* sobre determinada marca, de modo que quanto maior o percentual de *tweets* positivos em relação ao número de *tweets* publicados, maior seria a qualidade percebida.

Entretanto, analisando os resultados a partir da aplicação do modelo proposto, percebeu-se que os mesmos não representavam aquilo que os consumidores percebiam em relação à qualidade do produto, mas davam apenas ideia da opinião

geral dos consumidores em relação a essas marcas. Relembrando o que Aaker (1998) ensina, a qualidade do produto está associada a sete dimensões (Figura 5) – desempenho, características, conformidade com as especificações, confiabilidade, durabilidade, disponibilidade de serviços e forma e acabamento – ao passo que a qualidade do serviço está associada a outras oito – tangibilidade, confiabilidade, competência da equipe, atendimento, empatia, credibilidade, honestidade e cortesia.

Aaker (1998) também recorda que esse importante elemento da equidade da marca não está ligado, necessariamente, à qualidade real do produto, mas sim à maneira como essa qualidade é captada pelos consumidores. Dessa forma, a qualidade percebida deve ser descrita por meio dos conhecimentos e julgamentos dos próprios clientes dessas marcas.

Nam e Kannan (2014), ao classificar rótulos coletados da rede social *Delicious* para inferir sobre os tipos de associações que podem ser construídas com uma marca, perceberam que apenas 4,5% desses rótulos estavam ligados à avaliação do produto, sendo que as associações restantes correspondiam a atributos tangíveis e intangíveis do produto (5,1%), categorias de produto (6,9%), marcas propriamente dita (12,2%), nomes ou lugares (23,2%), outros substantivos (41,6%) e demais palavras (6,2%).

Na língua portuguesa, conforme definem Cunha e Cintra (2008), os adjetivos podem ser definidos como toda palavra que caracteriza um substantivo, indicando-lhes suas qualidades, defeitos, aspectos, estados ou aparências.

Com base nas premissas apresentadas, sugere-se que no modelo proposto dessa pesquisa os adjetivos relacionados a um produto ou serviço sejam utilizados para compreender melhor quais as principais qualidades atribuídas, pelos consumidores, às marcas. Dessa forma, realizando-se uma análise morfológica das palavras recuperadas a partir das publicações dos usuários no Twitter, podem-se isolar os adjetivos dessas sentenças e utilizá-los para análise da qualidade percebida.

#### 4.4.4 Sentimento percebido

No modelo proposto, apresenta-se a inclusão de um novo elemento para a análise da equidade da marca a partir das redes sociais, denominado sentimento percebido. Esse novo componente corresponde exatamente ao elemento qualidade percebida do modelo preliminar.

Deste modo, o sentimento percebido deve descrever o sentimento geral dos consumidores em relação à determinada marca, por meio da relação entre os *tweets* positivos publicados sobre a marca e o somatório de *tweets* positivos e negativos relativos à marca:

$$sp'(X) = \frac{pos(x)}{pos(x) + neg(x)} \quad (24)$$

Onde:

$sp'(x)$ : sentimento percebido normalizado de  $x$

$pos(x)$ : número de *tweets* positivos sobre a marca  $x$

$neg(x)$ : número de *tweets* negativos sobre a marca  $x$

De modo análogo ao modelo preliminar, o sentimento percebido normalizado foi ponderado pela métrica PageRank, que determina, recursivamente, o prestígio de um elemento da rede a partir do prestígio dos elementos com os quais está ligado, uma vez que essas ligações indicam um implícito transporte de autoridade entre esses elementos:

$$spp'(x) = \frac{\sum_{i=1}^I P(i) * pos(i)}{\sum_{i=1}^I P(i) * pos(i) + \sum_{j=1}^J P(j) * neg(j)} \quad (25)$$

$spp'(x)$ : sentimento percebido ponderado normalizado de  $x$

$I$ : número de elementos que emitiram opinião positiva sobre  $x$

$P(i)$ : PageRank do elemento emissor de opinião positiva

$pos(i)$ : número de mensagens positivas sobre  $x$  emitidas pelo nó  $i$

$J$ : número de elementos que omitiram opinião negativa

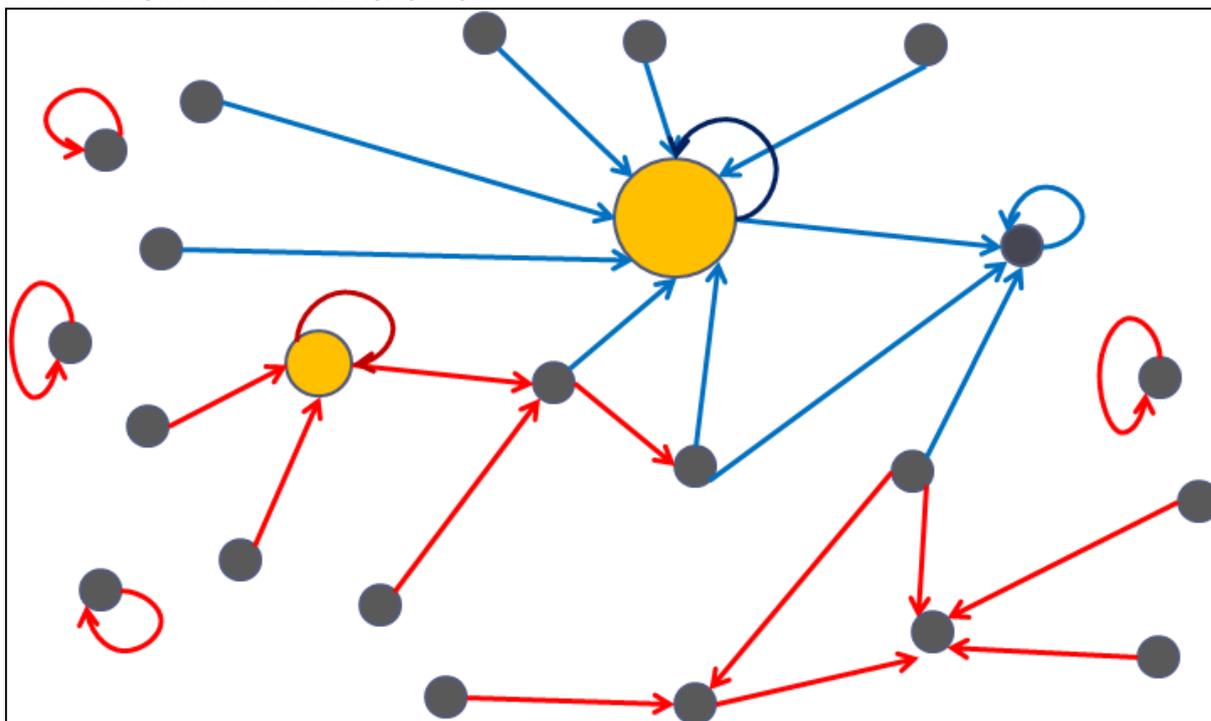
$P(j)$ : PageRank do elemento emissor de opinião negativa

$neg(j)$ : número de mensagens negativas sobre  $x$  emitidas pelo nó  $j$

Existe, entretanto, uma importante diferença entre a qualidade percebida do modelo preliminar e o sentimento percebido do modelo proposto. Neste, os usuários classificados como não-consumidores foram excluídos da análise, conforme

procedimento descrito na Subseção 3.4.2.2.6. Dessa forma, minimiza-se o impacto desses perfis na análise de sentimento, de modo que esta seja realizada a partir dos *tweets* publicados apenas por potenciais consumidores. O impacto dessa diferença pode ser mais bem explicado por meio da Figura 58.

**FIGURA 58 – PROPOSTA DO MODELO PARA ESTABELECIMENTO DO SENTIMENTO PERCEBIDO A PARTIR DE UMA REDE DE MENSAGENS**



FONTE: O AUTOR (2015)

Na Figura 58, o tamanho dos nós que representam os perfis é diretamente proporcional ao prestígio desses perfis na rede, com base na métrica PageRank; além disso, os nós em amarelo representam perfis não-consumidores. Isso significa dizer que os *tweets* publicados por esses perfis – destacados nas cores azul escuro e vermelho escuro – passam a não fazer mais parte da análise. Para o cálculo do sentimento percebido normalizado ( $sp'(x)$ ), essa alteração exerce um menor impacto, uma vez que a proporção de perfis consumidores é, geralmente, maior se comparada à proporção de perfis não-consumidores. Entretanto, quando se analisa o resultado dessa alteração sobre o cálculo do sentimento percebido ponderado normalizado ( $spp'(x)$ ), percebe-se que o impacto é maior, justamente pelo fato de que os perfis não-consumidores serem aqueles que, geralmente, possuíam maior prestígio na rede de publicações.

Relevante, porém, fazer uma observação: o sentimento associado às mensagens publicadas por perfis classificados como não-consumidores podem ainda sim influenciar o cálculo do sentimento percebido, caso os perfis classificados como consumidores realizem o *retweet* dessas mensagens. Em todo caso, o peso dessas mensagens ainda assim será menor no cálculo do sentimento percebido ponderado, uma vez que o sentimento atribuído a essas mensagens será ponderado pelo PageRank daqueles que fizeram o *retweet* e não pelo PageRank do perfil não-consumidor, emissor original da mensagem.

#### 4.4.5 Associações à marca

As associações à marca, assim como no modelo preliminar, são descritas mediante a frequência das palavras presentes nos *tweets* referentes às marcas analisadas. Essa representação é feita visualmente por meio de uma nuvem de palavras, na qual o tamanho das palavras é diretamente proporcional à frequência com que as mesmas aparecem nos *tweets*.

A eliminação de perfis não-consumidores, conforme descrito na Subseção 3.4.2.2.6, também foi incorporada a essa análise. Assim, as palavras presentes nos *tweets* publicados por perfis classificados como não-consumidores passam a não ser associadas à marca, uma vez que, assim como os demais elementos da equidade da marca, as associações devem representar apenas as percepções dos consumidores.

A seguir, na Seção 4.5, são apresentados os resultados da aplicação do modelo proposto sobre os dados coletados da rede social Twitter, referentes a quatro marcas de montadoras presentes no Brasil.

## 4.5 RESULTADOS DA APLICAÇÃO DO MODELO PROPOSTO

Após a execução dos passos de coleta e preparação dos dados referentes à etapa de teste, aplicou-se o modelo proposto, resultante de alterações apresentadas com base na análise crítica do modelo preliminar. Os resultados obtidos a partir da aplicação desse modelo são apresentados nessa seção.

O primeiro elemento a ser analisado sob a óptica do modelo proposto foi o conhecimento da marca. Os resultados obtidos para as variáveis conhecimento absoluto ( $c$ ) e conhecimento normalizada ( $c'$ ) são indicados na Tabela 17.

**TABELA 17 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PARA O CRITÉRIO CONHECIMENTO DA MARCA REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS PESQUISADAS**

x	followers	$c(x)$	$c'(x)$
Chevrolet	57187	57187	0,263439870
Fiat	92281	92281	0,425105262
Renault	83380	83380	0,384101567
Volkswagen	42034	42034	0,193635467

**FONTE: O AUTOR (2015)**

Conforme se percebe com base nos resultados apresentados, a Fiat é aquela que detém maior conhecimento dentre as marcas de montadoras analisadas, sendo seguida por aproximadamente 42,5% dos perfis que compõe a base da pesquisa, ficando à frente de suas concorrentes diretas Renault, Chevrolet e Volkswagen, respectivamente.

Para a adequação do conhecimento relativo da marca aos valores presentes na pirâmide do conhecimento de Aaker (1998), uma conversão dos valores contínuos para nominais foi proposta, conforme apresentado na Figura 56. Os valores obtidos por meio da aplicação dessa transformação são apresentados na Tabela 18.

**TABELA 18 – CONVERSÃO DOS VALORES OBTIDOS PARA CONHECIMENTO RELATIVO DAS MARCAS PARA OS NÍVEIS DE CONHECIMENTO DA MARCA**

x	$c'(x)$	Nível de conhecimento
Chevrolet	0,263439870	Lembrança da marca
Fiat	0,425105262	<i>Top of mind</i>
Renault	0,384101567	<i>Top of mind</i>
Volkswagen	0,193635467	Reconhecimento da marca

**FONTE: O AUTOR (2015)**

Após a conversão, pode-se observar que duas das quatro marcas – Fiat e Renault – encontram-se no mesmo nível de conhecimento, sendo classificadas como *Top of Mind*. Segundo a proposta original de Aaker (1998), essas são as primeiras marcas lembradas pelo consumidor quando os mesmos são estimulados a pensar em

uma categoria específica de produtos. A marca Chevrolet, por sua vez, foi classificada como lembrada, enquanto que a marca Volkswagen foi rotulada apenas como reconhecida.

A baixa classificação da marca Volkswagen em relação ao reconhecimento é consequência do baixo número de seguidores dessa marca na rede social Twitter, em relação às suas concorrentes diretas. Isso pode ser resultado das diferentes estratégias de marketing entre as marcas ou até mesmo um enfoque em diferentes redes sociais.

Outro fator que pode influenciar no menor número de seguidores da marca Volkswagen em comparação as suas concorrentes: enquanto que parece haver certo padrão relacionado ao nome do perfil de suas concorrentes (@fiatbr, @renaultbrasil e @chevroletbrasil), o perfil da Volkswagen no Brasil é @carromesmo, o que pode criar uma dificuldade maior para ser encontrado pelos fãs da marca.

Apesar de ser a última colocada quando o fator conhecimento da marca é considerado, a Volkswagen supera todos seus concorrentes no segundo elemento concernente à equidade das marcas nas redes sociais: a lealdade à marca. Os valores obtidos para lealdade absoluta e lealdade normalizada são apresentados na Tabela 19.

**TABELA 19 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PARA O CRITÉRIO LEALDADE À MARCA REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS PESQUISADAS**

x	I(x)	I'(x)
Chevrolet	24956	0,436392886
Fiat	66764	0,723485875
Renault	53277	0,638966179
Volkswagen	32167	0,765261455

**FONTE: O AUTOR (2015)**

Apesar de a Volkswagen apresentar um menor valor absoluto de lealdade, quando esse elemento é avaliado considerando-se o valor total de seguidores de cada marca, a montadora de origem alemã obtém um índice de aproximadamente 0,76, o que significa que apenas 24% de seus seguidores segue pelo menos uma das demais marcas presentes na pesquisa. Na outra extremidade do ranking de lealdade ficou a Chevrolet, já que 56,4% de seus seguidores também seguem pelo menos um de seus concorrentes presentes na pesquisa.

Conforme propõe o modelo, foi realizada a transformação dos valores contínuos de lealdade em valores nominais, seguindo as regras apresentadas na Figura 57. Os resultados obtidos como decorrência desse processo são mostrados na Tabela 20.

**TABELA 20 – CONVERSÃO DOS VALORES OBTIDOS PARA LEALDADE RELATIVA DAS MARCAS PARA OS NÍVEIS DE LEALDADE DA MARCA**

x	$l'(x)$	Nível de lealdade
Chevrolet	0,436392886	Satisfeito, mas menos vulnerável à mudança
Fiat	0,723485875	Gosta da marca, considera-a amiga
Renault	0,638966179	Gosta da marca, considera-a amiga
Volkswagen	0,765261455	Gosta da marca, considera-a amiga

**FONTE: O AUTOR (2015)**

Assim, três das quatro marcas presentes na pesquisa – Volkswagen, Fiat e Renault – foram associadas ao segundo nível de lealdade da pirâmide de Aaker, que representa os consumidores que gostam da marca e a consideram amiga. Por outro lado, a marca Chevrolet foi associada ao terceiro nível da pirâmide de lealdade, que descreve os consumidores satisfeitos e pouco vulneráveis a mudanças. Com base nos resultados da pesquisa, nenhuma das quatro marcas de montadoras atingiu o maior nível da pirâmide de lealdade, que representa os consumidores comprometidos. Para atingir esse nível, uma marca precisaria ter, no máximo, 20% de seguidores que também seguissem uma marca concorrente.

Na Tabela 21 são apresentados os resultados obtidos para o novo critério proposto para análise da equidade das marcas nas redes sociais: o sentimento percebido. Esse critério foi adotado para contemplar o sentimento geral dos consumidores em relação às marcas, ocupando o papel do critério qualidade percebida, presente no modelo preliminar.

**TABELA 21 – RESULTADOS OBTIDOS COM A APLICAÇÃO DO MODELO PARA O CRITÉRIO SENTIMENTO PERCEBIDO REFERENTES A CADA UMA DAS QUATRO MARCAS AUTOMOBILÍSTICAS PESQUISADAS**

x	sp'(x)	spp'(x)
Chevrolet	0,721817	0,727069
Fiat	0,480023	0,463272
Renault	0,688620	0,673222
Volkswagen	0,782662	0,788493

FONTE: O AUTOR (2015)

Com base nos dados apresentados na Tabela 21, a marca Volkswagen foi a que obteve um maior índice para o critério sentimento percebido, tanto na forma ponderada quanto na não ponderada. Assim, apesar de ser a última classificada no critério conhecimento da marca, a Volkswagen superou suas concorrentes nos outros dois critérios analisados até aqui: lealdade da marca e sentimento percebido.

Por outro lado, a marca Fiat, que havia se colocado bem perante suas concorrentes nos dois critérios analisados anteriormente, foi a que alcançou o menor índice de sentimento percebido, também se consideradas as duas implementações – ponderada e não ponderada – desse elemento.

Como pode ser observado, verificou-se uma menor variação dos valores de sentimento percebido normalizado (sp') e sentimento percebido ponderado normalizado (spp'), se comparado do modelo preliminar. Ao contrário do antigo modelo, foram considerados como entrada para essa análise apenas os *tweets* cujos perfis emitentes foram classificados como consumidores. Portanto, a exclusão dos perfis não-consumidores do processo de análise fez com que a qualidade percebida sofresse uma menor interferência do prestígio – com base no cálculo do PageRank – desses elementos na rede que, em média, é maior do que o prestígio dos elementos consumidores.

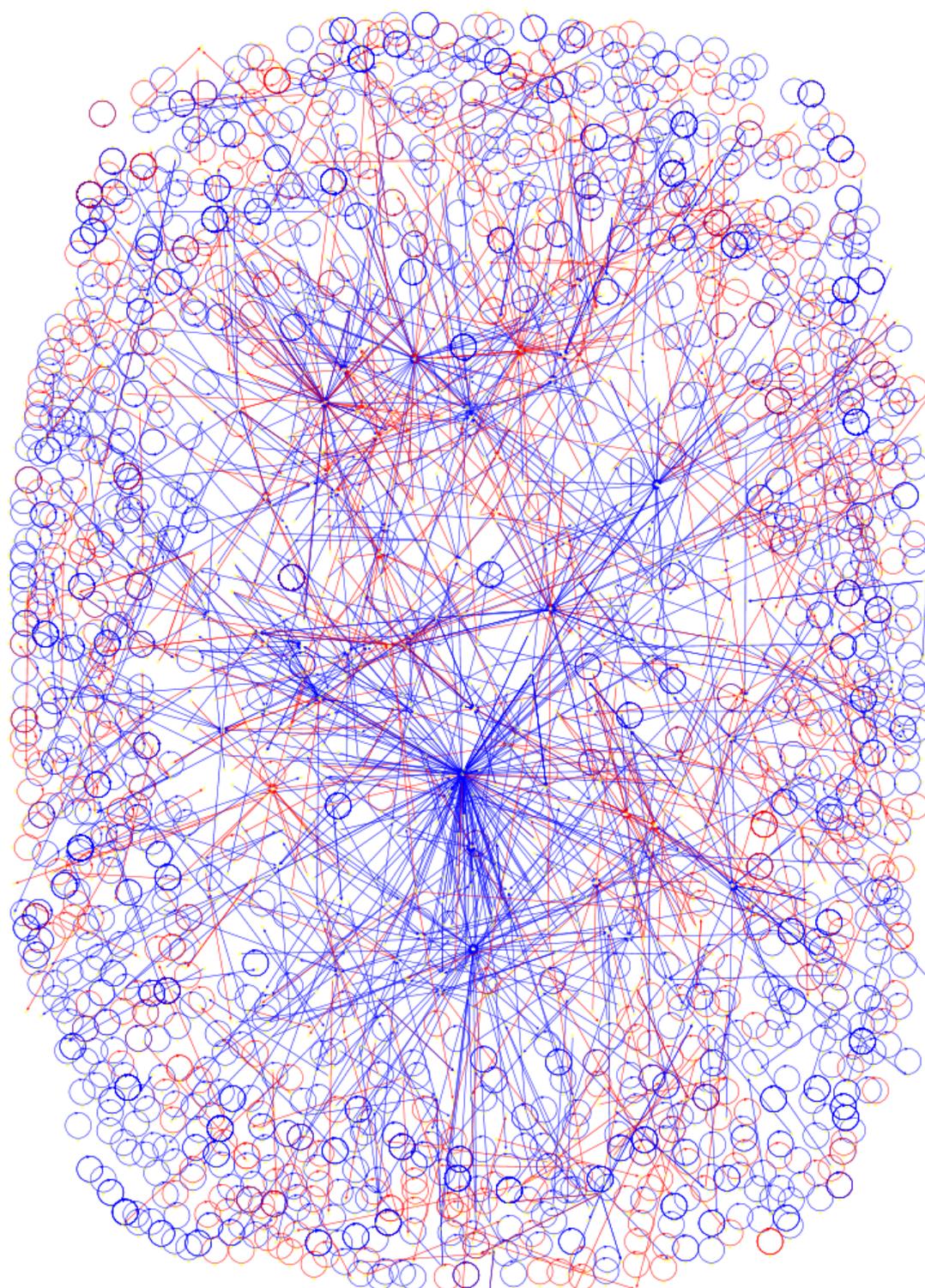
A rede de *tweets* referentes a marca Chevrolet, publicados por perfis classificados como consumidores, é mostrada na Figura 59. Nessa rede, os auto-relacionamentos representam as publicações dos usuários consumidores, enquanto que as ligações lineares entre diferentes pontos representam as relações de *retweet* ou referências a outros perfis.

Cabe salientar que, apesar da eliminação dos *tweets* publicados por perfis classificados como não consumidores, esses nós ainda podem aparecer na rede, caso algum usuário classificado como consumidor *retweet* uma publicação de um perfil

classificado como não consumidor ou ainda faça uma referência a um perfil desse tipo. Entretanto, esses perfis não consumidores que aparecem na rede não interferem diretamente no sentimento percebido ponderado, já que as mensagens publicadas por eles não foram consideradas na análise.

Ainda considerando a Figura 59, pode ser observado um predomínio de *tweets* relacionado a marca Chevrolet classificados com sentimento predominantemente positivo – representados em azul – em relação àqueles classificados com sentimento predominantemente negativo – representados em vermelho. Dos 3.699 *tweets* relacionados à marca Chevrolet, aproximadamente 72% (2.670) foram classificados como positivos, enquanto que 28% (1.029) foram classificados como negativos.

**FIGURA 59 – REDE DE TWEETS RELACIONADOS À MARCA CHEVROLET CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS**

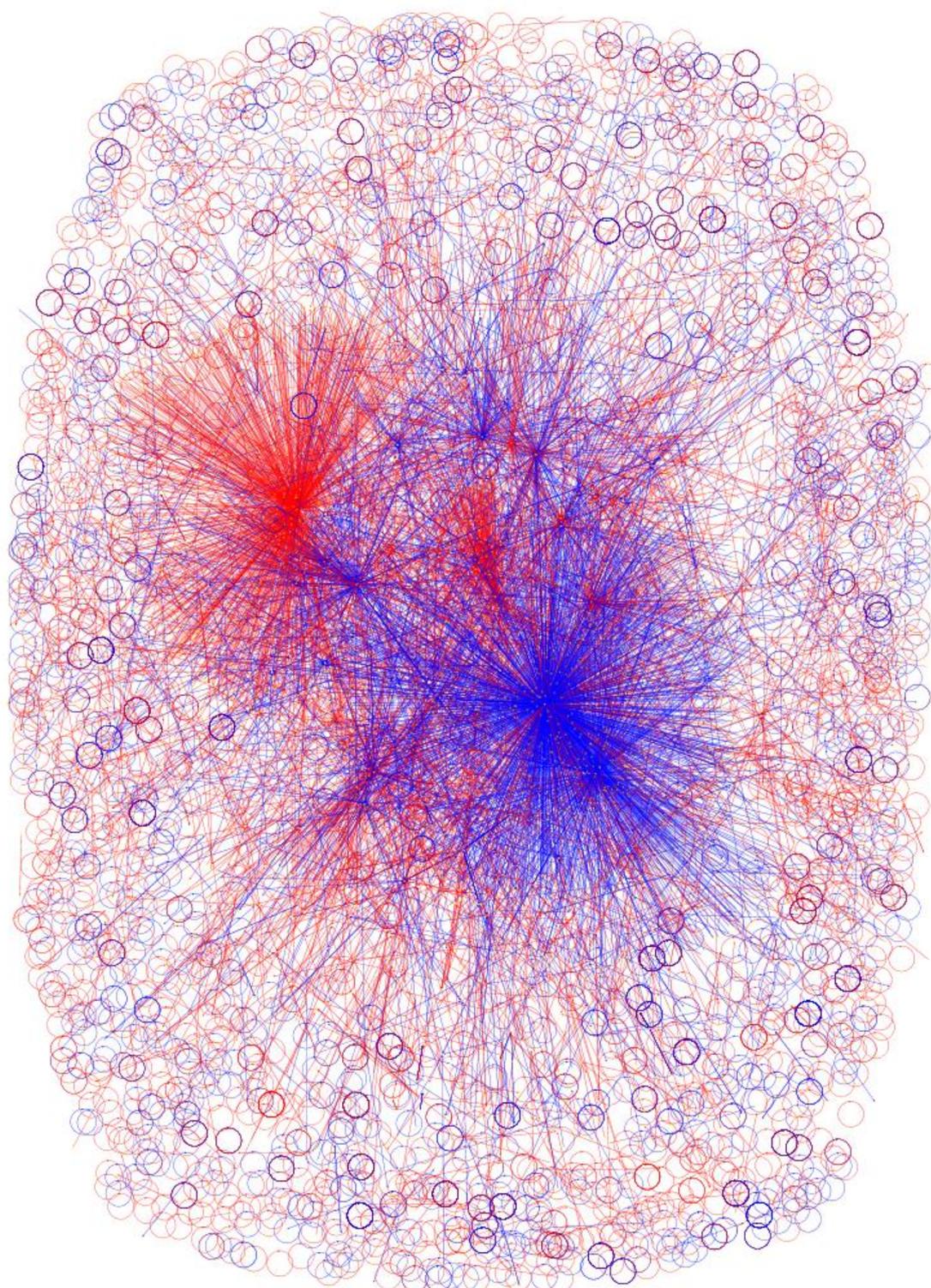


Created with NodeXL (<http://nodexl.codeplex.com>)

**FONTE: O AUTOR (2015)**

Na Figura 60, por sua vez, é apresentada a rede de *tweets* relacionados à marca Fiat, publicados por perfis classificados como consumidores. Ao contrário da Figura 59, observa-se na Figura 60 um leve predomínio de publicações negativas. Dos 10.387 *tweets* analisados sobre a marca Fiat, 52% (5.401) foram classificados como negativos contra 48% positivos (4.986). Pode ser observada também uma maior densidade dessa rede em relação à rede de *tweets* da marca Chevrolet, o que pode ser explicado pelo maior volume de *tweets* referentes à marca Fiat.

FIGURA 60 – REDE DE TWEETS RELACIONADOS À MARCA FIAT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



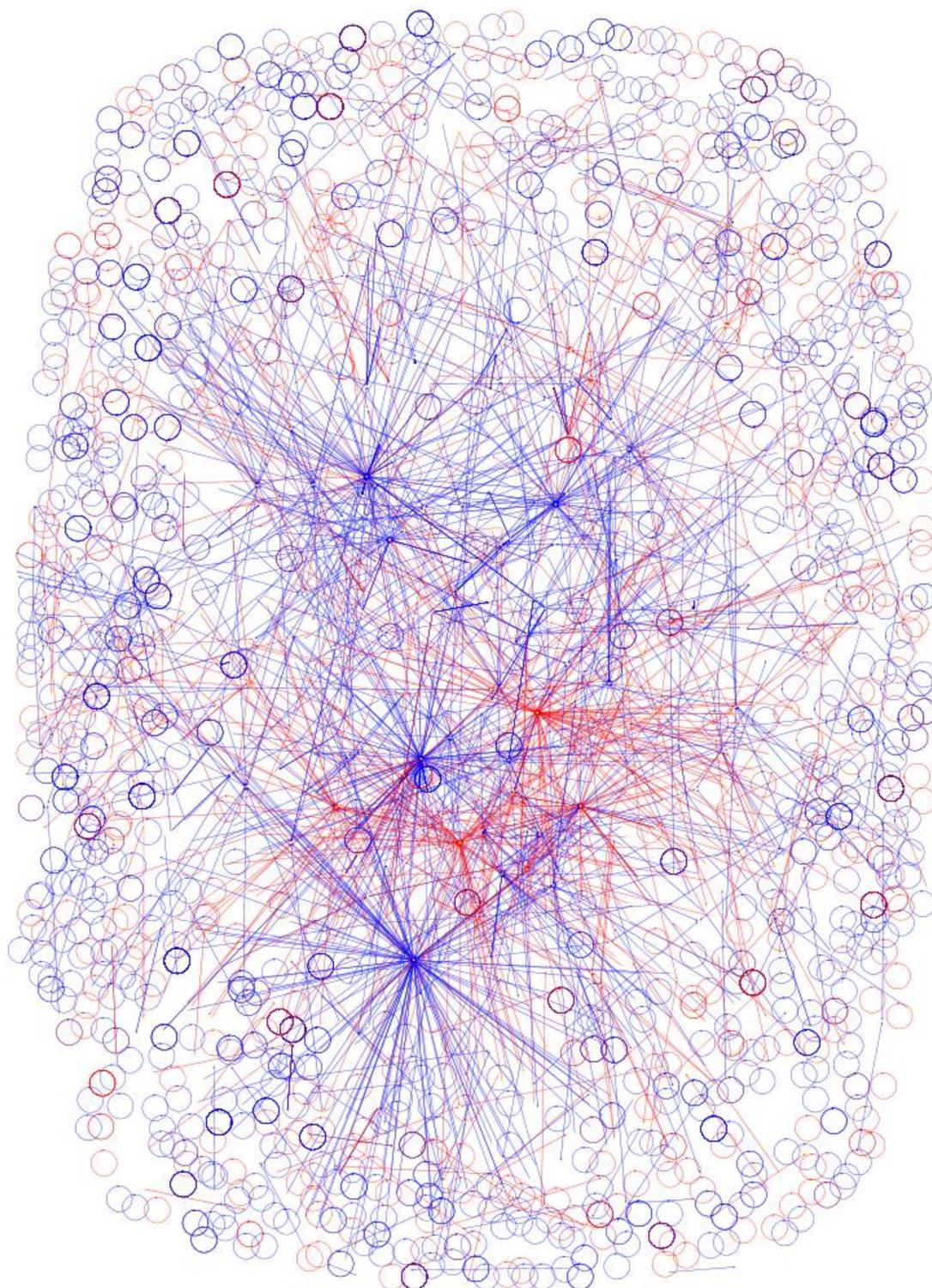
Created with NodeXL (<http://nodexl.codeplex.com>)

FONTE: O AUTOR (2015)

A rede de *tweets* relacionados à marca Renault, por sua vez, é a menos densa dentre as todas as redes analisadas, dado ao menor volume de *tweets* publicados

sobre a marca por perfis consumidores, totalizando 3.796 publicações. Essa rede é apresentada na Figura 61.

**FIGURA 61 – REDE DE TWEETS RELACIONADOS À MARCA RENAULT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS**

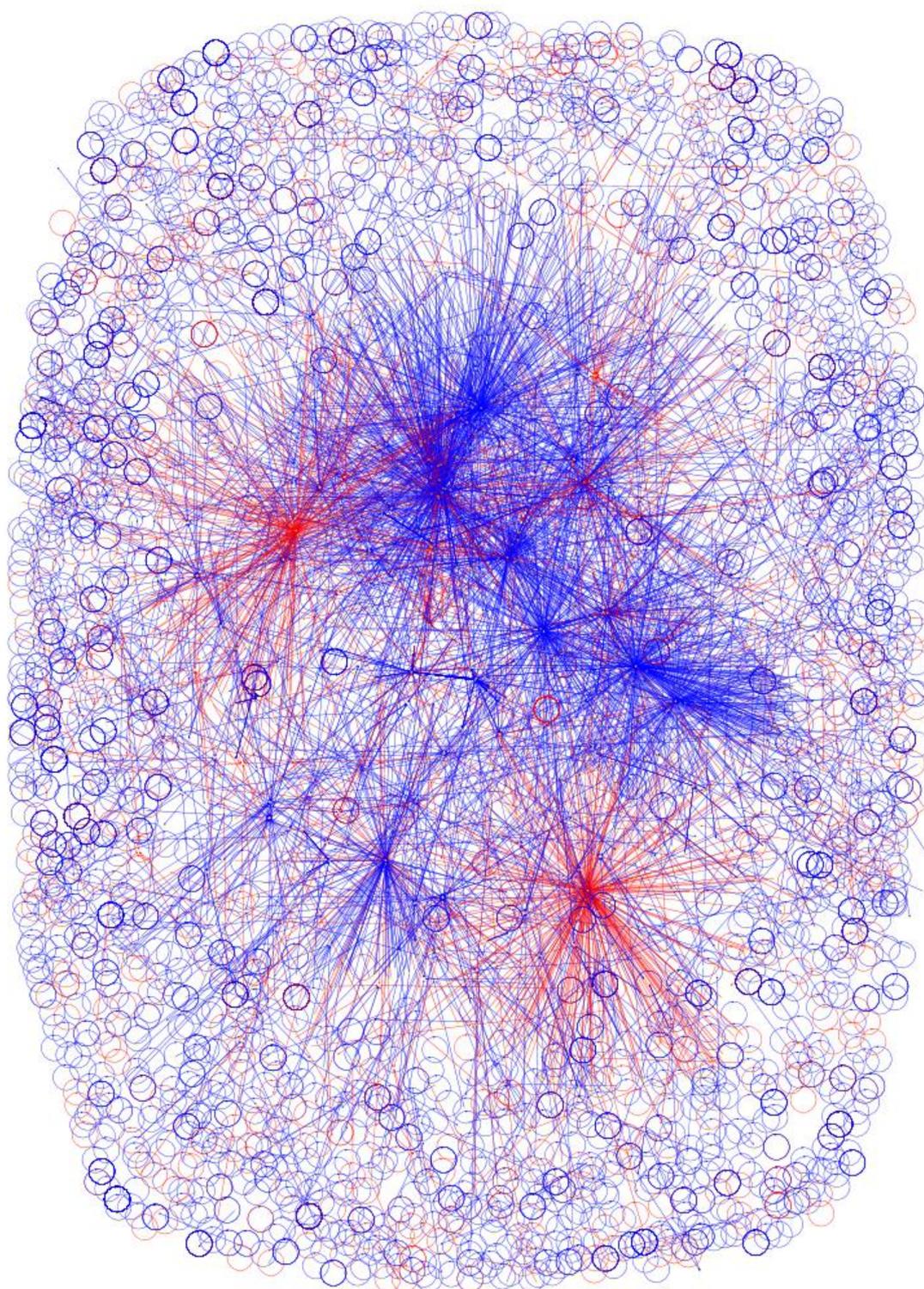


Created with NodeXL (<http://nodexl.codeplex.com>)

**FONTE: O AUTOR (2015)**

Observa-se na rede mostrada na Figura 61 um predomínio maior dos *tweets* positivos (69%, que corresponde a 2.614 publicações) em relação àqueles classificados como negativos (31%, que corresponde a 1.182 publicações). Por fim, com a maior proporção de publicações positivas em relação às negativas, a rede de *tweets* relacionados à marca Volkswagen é apresentada na Figura 62. Atingindo o segundo lugar em volume, com 8190 publicações, das quais 78% (6.410) são positivas e 22% (1.780), negativas.

**FIGURA 62 – REDE DE TWEETS RELACIONADOS À MARCA VOLKSWAGEN CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS**



Created with NodeXL (<http://nodexl.codeplex.com>)

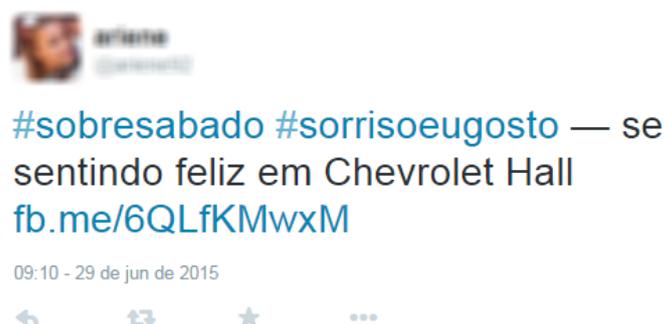
**FONTE: O AUTOR (2015)**

Assim como no modelo preliminar, as associações às marcas também foram analisadas. Mais uma vez, a diferença fundamental entre os dois modelos é em



analisados referentes à marca. Em segundo lugar vem a palavra “hall”, que aparece 413 vezes, principalmente relacionadas a *tweets* que citam as casas de show Chevrolet Hall, presentes tanto em Belo Horizonte quanto em Recife, como no exemplo apresentado na Figura 64.

**FIGURA 64 – EXEMPLO DE TWEET RELACIONADO À PALAVRA “HALL”**



**FONTE: COM BASE EM DADOS DO TWITTER (2015)**

Conforme afirmam Ashley e O’Hara (2001), essa prática de associar nomes de marcas à estabelecimentos, projetos ou eventos é conhecida como *naming rights*, sendo um dos principais veículos utilizados pelas empresas para aumentar a consciência e o reconhecimento de uma marca. A análise dos *tweets* coletados indica que essa é uma prática bastante eficaz, já que a segunda palavra mais presente nos *tweets* relacionados à marca Chevrolet faz referência exatamente a um estabelecimento batizado com o nome da marca.

As palavras “silverado” “cruze”, “camaro” e “onix”, que representam modelos de carros relacionados à marca Chevrolet, também aparecem em destaque na nuvem de palavras da marca, respectivamente nas 3ª, 4ª, 11ª e 12ª posições no ranking de associações.

Na Figura 65, por sua vez, são apresentadas as palavras mais associadas a *tweets* positivos sobre a marca Chevrolet.

FIGURA 65 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS POSITIVOS SOBRE A MARCA CHEVROLET CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

Mais uma vez, destacam-se as palavras “chevrolet”, “silverado”, “cruze”, “hall”, e “camaro”, ocupando, respectivamente, a 1ª, 2ª, 3ª, 4ª e 10ª posições no ranking de associações. Entretanto, observa-se uma importante variação em relação ao ranking geral: a palavra “onix”, que no ranking geral ocupava a 12ª posição, ocupa apenas a 98ª no ranking contendo apenas as associações relacionadas a *tweets* classificados como positivos, indicando que a maioria das associações com a palavra “onix”



FIGURA 67 – EXEMPLOS DE TWEETS NEGATIVOS CONTENDO A PALAVRA “ONIX”




**Mini Mini Anne**  
 @anna\_anna  
 PESSOAL DE POA:  
 UM AMIGO FOI ASSALTADO AI E  
 LEVARAM O CARRO DELE! UM  
 CHEVROLET ONIX PRETO, PLACA [redacted]  
 [redacted]  
 SE TIVEREM INFOS LIGUEM PRO 190

FONTE: COM BASE EM DADOS DO TWITTER (2015)

Outra característica das publicações relacionadas a roubos de veículos é o número elevado de *retweets*, o que acaba aumentando o impacto desse tipo de postagem na análise.

As palavras associadas à montadora Fiat também foram analisadas. A nuvem das palavras mais frequentes nos *tweets* relacionados à marca é apresentada na Figura 68.



FIGURA 69 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS POSITIVOS SOBRE A MARCA FIAT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

Apesar de o modelo “elba” aparecer em destaque *tweets* positivos relacionados à marca Fiat, com 895 ocorrências, há uma predominância dessa palavra nos *tweets* negativos, onde 1.405 ocorrências foram encontradas. O modelo “strada”, por sua vez, apesar de, em números absolutos, aparecer menos que o modelo “elba” nos *tweets* positivos, supera esse modelo se considerada a proporção entre positivos e negativos, já que aparece 627 vezes em *tweets* positivos e 110 em negativos.

Essa grande diferença entre ocorrências em *tweets* positivos e negativos é alavancada pelo grande número de *retweets* de uma publicação realizada pelo perfil oficial da marca Fiat, apresentado na Figura 70.

FIGURA 70 – EXEMPLOS DE TWEET POSITIVO CONTENDO A PALAVRA “STRADA”



FONTE: COM BASE EM DADOS DO TWITTER (2015)

Destaca-se, mais uma vez, que apesar de os *tweets* publicados por perfis classificados como não-consumidores (como o exemplo da Figura 70) não influenciarem diretamente na análise de associações, eles impactam indiretamente, uma vez que podem ser propagados na rede por usuários cujos perfis foram classificados como consumidores.

FIGURA 71 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE A MARCA FIAT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

A predominância das palavras “elba” e “collor” em *tweets* negativos relacionados à marca Fiat é decorrente de críticas ou sátiras realizadas por usuários do Twitter como reação à notícia de que a polícia federal havia apreendido carros de Collor na segunda fase da operação lava jato. Alguns exemplos dessas publicações podem ser visualizados na Figura 72.

FIGURA 72 – EXEMPLOS DE TWEETS NEGATIVOS CONTENDO AS PALAVRAS “ELBA” E “COLLOR”



FONTE: COM BASE EM DADOS DO TWITTER (2015)

Outros usuários, por sua vez, comparam os cenários de *impeachment* do ex-presidente Collor com o cenário enfrentado pelo governo Dilma, como exemplificado na Figura 73.

**FIGURA 73 – EXEMPLOS DE TWEETS NEGATIVOS, ASSOCIADOS A PROCESSOS DE IMPEACHMENT, CONTENDO A PALAVRA “FIAT”**



**FONTE: COM BASE EM DADOS DO TWITTER (2015)**

Também predominante em *tweets* negativos relacionados à marca Fiat, o modelo “uno” aparece geralmente em *tweets* relacionados à roubos ou acidentes, como nos exemplos apresentados na Figura 74.

**FIGURA 74 – EXEMPLOS DE TWEETS NEGATIVOS, ASSOCIADOS A ROUBOS OU ACIDENTES, CONTENDO A PALAVRA “UNO”**



#alertarobertatrindade

O Fiat Uno prata placa [REDACTED] está abandonado no Morro da Jaqueira, no bairro Paraíso,... [fb.me/1X0KVoylc](https://www.facebook.com/1X0KVoylc)

14:03 - 28 de jun de 2015



#eSeu Fiat Uno prata, [REDACTED] abandonado na altura do nº280 da Estrada de Botafogo, numa vila, em Costa Barros, perto a UPA.

RETWEET  
1

CURTIU  
1



04:10 - 7 de jul de 2015



@radiobandnewsfm motoqueiro bate na traseira do fiat uno em piratininga niterói



03:35 - 29 de jun de 2015



Na manhã desta terça-feira, 30, um Fiat Uno de placa [REDACTED] caiu em uma ribanceira na BR-101, na conhecida... [fb.me/4FJjRdqX4](https://www.facebook.com/4FJjRdqX4)



Santo Antônio de Jesus: Veículo cai em ribanceira na Cur...

blogdomartins.com.br

05:37 - 30 de jun de 2015



FIGURA 76 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS POSITIVOS SOBRE A MARCA RENAULT CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

Nos *tweets* positivos apresentados na Figura 77, pode-se observar um *feedback* positivo dos usuários classificados como consumidores em relação ao comercial da Renault Duster protagonizado por uma simpática senhorinha. A predominância dessas palavras em *tweets* positivos relacionados à marca é também alavancada pelo compartilhamento de notícias relacionadas ao novo modelo de pick-up lançada pela marca, a Duster Oroch.

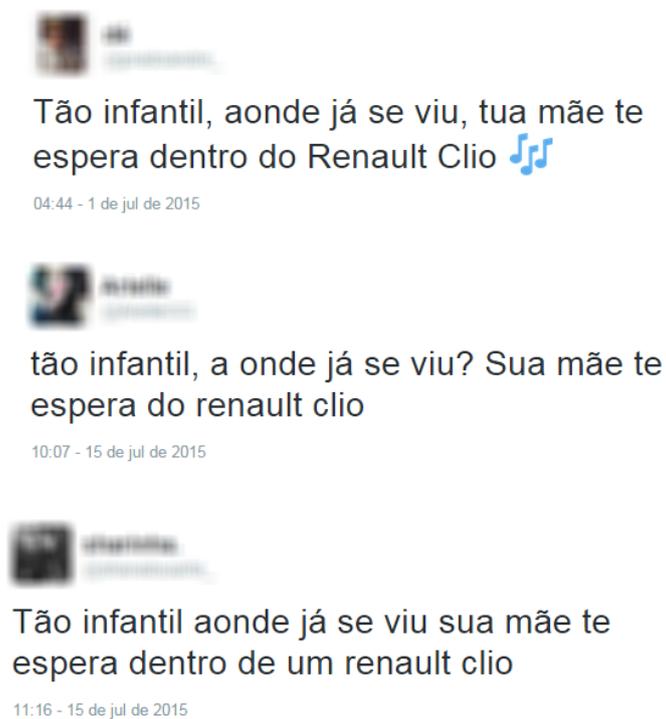
FIGURA 77 – EXEMPLOS DE TWEETS POSITIVOS CONTENDO AS PALAVRAS “RENAULT” E “DUSTER”



FONTE: COM BASE EM DADOS DO TWITTER (2015)

Por outro lado, nos *tweets* negativos, há a predominância das palavras “clio” e “placa”, como mostrado na Figura 78.



**FIGURA 79 – EXEMPLOS DE TWEETS NEGATIVOS CONTENDO A PALAVRA “CLIO”**

**FONTE: COM BASE EM DADOS DO TWITTER (2015)**

A grande ocorrência da palavra “placa”, mais uma vez, demonstra a relação de *tweets* negativos com roubos de veículos, como exemplificam as publicações que compõem a Figura 80.

FIGURA 80 – EXEMPLOS DE TWEETS NEGATIVOS CONTENDO A PALAVRA “PLACA”



FONTE: COM BASE EM DADOS DO TWITTER (2015)

Por fim, as associações referentes à marca Volkswagen são apresentadas por meio da nuvem de palavras que formam a Figura 81.

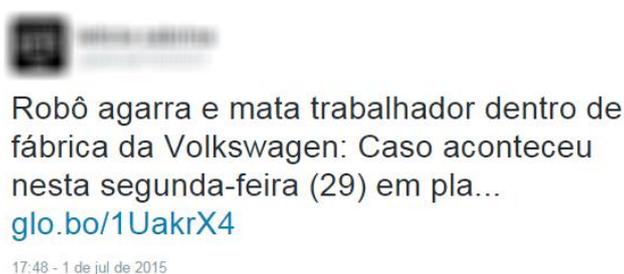
FIGURA 81 – NUVEM DE PALAVRAS RELACIONADAS À MARCA VOLKSWAGEN CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

As palavras que aparecem em destaque na nuvem da Figura 81 são referentes a uma notícia bastante difundida pelos usuários da rede social Twitter durante o período de análise da pesquisa. A notícia veiculada nos principais meios de comunicação em 1º de julho de 2015, intitulada “Robô mata funcionário da Volkswagen na Alemanha” (EPOCA, 2015) exerceu grande influência nas associações referentes à marca, sendo responsável pelo destaque das palavras “robô”, “mata”, “fábrica” e “funcionário” e “alemã”; a palavra “trabalhador”, presente em algumas variações da notícia, também aparece em destaque. Alguns tweets sobre a notícia podem ser observados na Figura 82.

FIGURA 82 – EXEMPLOS DE TWEETS RELACIONADOS AO ACIDENTE NA FÁBRICA DA VOLKSWAGEN NA ALEMANHA



FONTE: COM BASE EM DADOS DO TWITTER (2015)

Uma vez que a notícia foi classificada como positiva na maioria de suas variações, as palavras a ela relacionadas também predominam na nuvem de palavras

referentes aos *tweets* com sentimento predominantemente positivo, como pode ser visualizado na Figura 83.

FIGURA 83 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS POSITIVOS SOBRE A MARCA VOLKSWAGEN CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

A classificação da notícia como sendo positiva se dá principalmente pela ocorrência da palavra “mata” e suas variações – como “matar” e “matando” – no conjunto de *tweets* positivos que constituem a base utilizada para a classificação de sentimento. Alguns desses tweets, como “*se curtiu o vídeo dá like, se não, se mata! :D*”, “*@jamiauu hahahahaha, se não me matou com o gato, não mata mais :P*”, “*Que*

vontadezinha de mata alguém :)", "@gabibianco isso, me mata de inveja. :P", "gente n é ciumes só vontade de matar :))", "vontade de matar esse homem :)))" utilizam a palavra matar e suas variações no sentido figurado, o que acaba interferindo no processo de classificação.

Na nuvem de associações relacionadas aos tweets negativos da marca Volkswagen, apresentada por meio da Figura 84, destacam-se as palavras “gol”, “audi”, “bentley”, “dona”, “porsche”, “bugatti”, “ducati” e “lamborghini”.

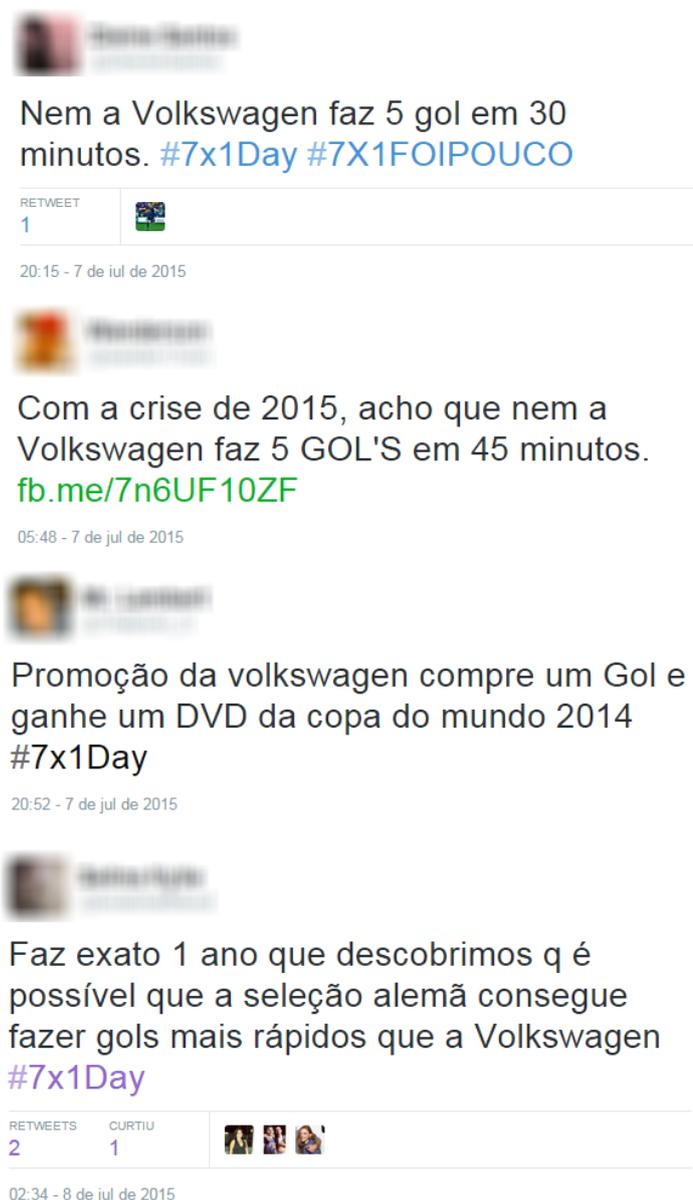
FIGURA 84 – NUVEM DE PALAVRAS ASSOCIADAS AOS TWEETS NEGATIVOS SOBRE A MARCA VOLKSWAGEN CONSTRUÍDA A PARTIR DAS PUBLICAÇÕES COLETADAS



FONTE: O AUTOR (2015)

Devido ao período em que a coleta dos *tweets* foi realizada, a palavra “gol”, referente ao modelo mais popular da marca no Brasil, aparece principalmente em publicações relembrando o aniversário de um ano da histórica goleada sofrida pelo Brasil na Copa de 2014, como pode ser verificado na Figura 85.

**FIGURA 85 – EXEMPLOS DE TWEETS COM A PALAVRA “GOL”**



**FONTE: COM BASE EM DADOS DO TWITTER (2015)**

As demais palavras destacadas estão relacionadas a uma curiosidade sobre a marca, bastante replicada pelos usuários da rede, como mostra a Figura 86.

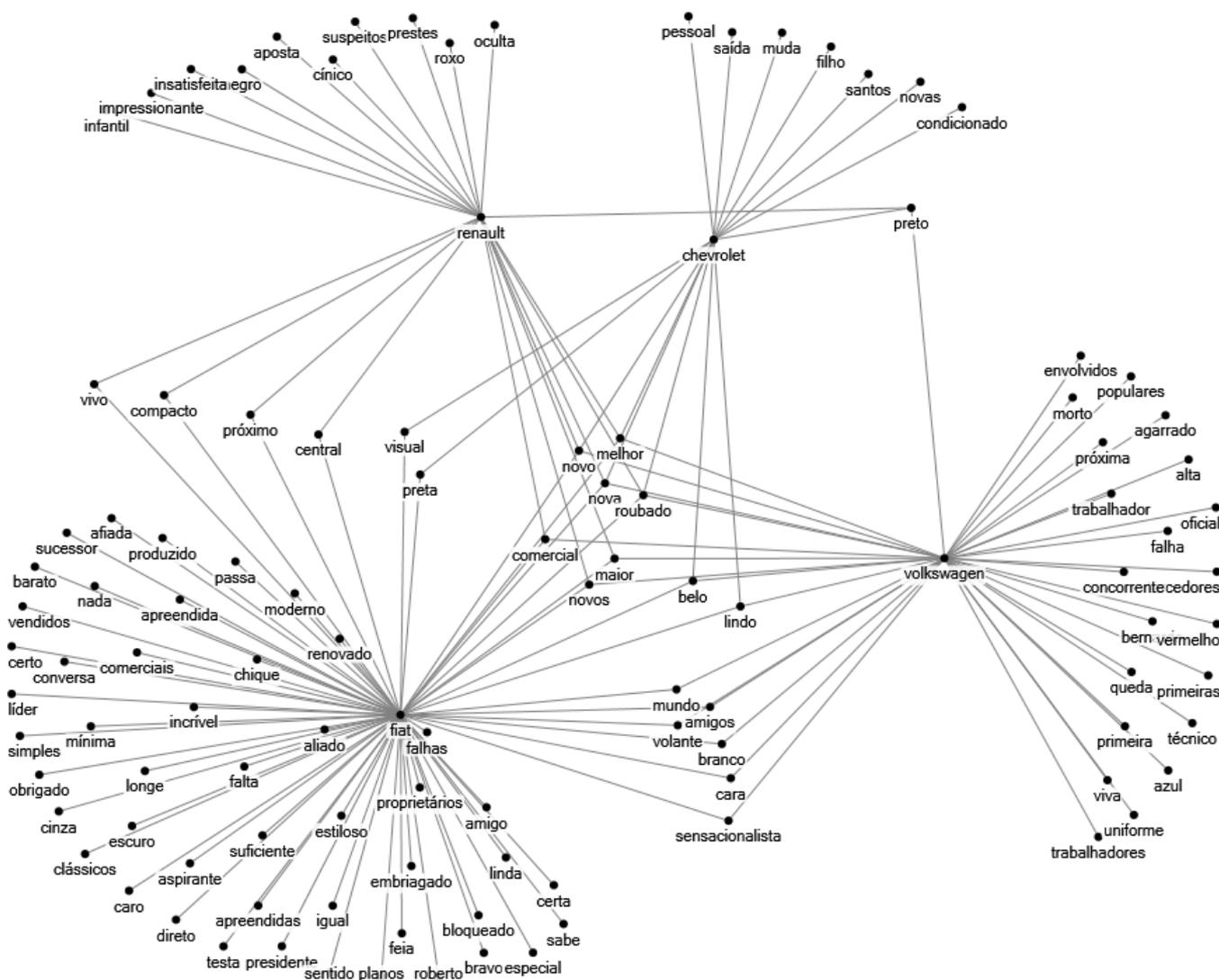
**FIGURA 86 – EXEMPLOS DE TWEETS PUBLICADOS SOBRE UMA CURIOSIDADE SOBRE A MARCA VOLKSWAGEN**



**FONTE: COM BASE EM DADOS DO TWITTER (2015)**

O último elemento da equidade das marcas nas redes sociais analisado foi a qualidade percebida. Diferentemente do modelo preliminar, a qualidade percebida é analisada a partir dos principais adjetivos presentes nos *tweets* sobre as marcas analisadas. A partir deles, busca-se identificar os principais atributos de qualidade associados às marcas pelos consumidores. Como base para essa análise, uma rede mostrando os adjetivos mais relacionados às marcas é mostrado na Figura 87. Para melhor visualização da rede, foram mostrados os adjetivos com, pelo menos, 20 ocorrências.

FIGURA 87 – REDE DE PALAVRAS OBTIDA PELA APLICAÇÃO DO MODELO PROPOSTO PARA O CRITÉRIO QUALIDADE PERCEBIDA, EM QUE SÃO APRESENTADOS OS ADJETIVOS MAIS ASSOCIADOS ÀS MARCAS PESQUISADAS



FONTE: O AUTOR (2015)

Dentre os atributos que possuem maior relação com a qualidade percebida pelos consumidores em relação às marcas, destacam-se “novo”, “nova” e “melhor”, que foram associados pelos consumidores às quatro marcas pesquisadas. O atributo “belo”, por sua vez, foi associado às marcas Chevrolet, Fiat e Volkswagen, mas não à marca Renault, assim como acontece com “lindo”. Já o adjetivo “compacto” aparece relacionado às marcas Fiat e Renault, mas não às marcas Chevrolet e Volkswagen.

Os demais atributos relacionados à qualidade aparecem associados a apenas uma das marcas, com destaques para “líder”, “barato”, “chique”, “moderno”, “estiloso”

e “caro”, associados à marca Fiat; “impressionantes” associado à Renault; e “populares” associado à Volkswagen.

#### 4.6 CONSIDERAÇÕES SOBRE O MODELO PROPOSTO

O modelo proposto para análise da equidade das marcas nas redes sociais foi construído a partir de alterações resultantes da análise e avaliação crítica dos resultados obtidos com a aplicação do modelo preliminar. Destacam-se, dentre essas alterações: a) a discretização dos elementos conhecimento da marca e lealdade a marca, para corresponder às pirâmides propostas por Aaker (1998); b) a inclusão de um novo elemento na equidade das marcas nas redes sociais: o sentimento percebido; c) a alteração do entendimento em relação ao elemento qualidade percebida; d) a eliminação dos perfis classificados como não consumidores na análise do sentimento percebido, qualidade percebida e associações às marcas.

A transformação dos resultados referentes aos elementos conhecimento da marca e lealdade à marca de contínuos para nominais permite uma melhor compreensão dos resultados pelos profissionais de marketing. Para eles, talvez seja mais valiosa a informação de que a marca foi classificada como *Top of Mind* do que se fosse transmitida a eles a informação de que a marca obteve o índice de 0,80 no critério de lealdade. Raciocínio análogo pode ser aplicado ao elemento conhecimento da marca.

Assim, acredita-se que a adequação dos valores obtidos nesses critérios para os níveis de conhecimento e lealdade propostos por Aaker (1998) seja uma das contribuições desse trabalho, já que nos modelos operacionais tradicionais para avaliação da equidade da marca, como os trabalhos de Yoo e Donthu (2001) e Pappu, Quester e Cooksey (2005), essa transformação não acontece, o que pode tornar difícil a compreensão por parte dos profissionais de marketing.

A proposta de um novo elemento para análise da equidade das marcas nas redes sociais é outra modificação incluída no modelo. O sentimento percebido busca mensurar a opinião geral dos consumidores em relação à marca. A partir da definição de equidade da marca proposta por Keller e Machado (2006, p.30), de que o “*brand equity* está relacionado ao fato de se obterem com uma marca resultados diferentes daqueles que se obteriam se o mesmo produto ou serviço não fosse identificado por

aquela marca” e da afirmação de Kotler e Keller (2006, p. 270), de que o “*brand equity* é o valor agregado atribuído a produtos e serviços [...] que pode se refletir no modo como os consumidores pensam, sentem e agem em relação à marca”, pode-se inferir que o sentimento associado pelos consumidores à marca contribui para a construção da equidade, à medida em que uma marca com sentimento positivo associado pode obter melhores resultados do que marcas acompanhadas de opiniões negativas.

A qualidade percebida, que no modelo preliminar buscava descrever exatamente esse sentimento associado às marcas, passou, no modelo proposto, a tentar compreender os principais predicados associados a essas marcas. Acredita-se que os resultados obtidos a partir da aplicação desse conceito foram satisfatórios, à medida em que, a partir desses predicados, pode-se identificar alguns elementos das dimensões da qualidade do produto ou serviço levantados por Aaker (1998).

A obtenção desses predicados, porém, foi realizada com base em uma análise morfológica das palavras presentes nas publicações coletadas sobre as marcas na rede social Twitter. Relembrando mais uma vez Cunha e Cintra (2008), os adjetivos podem ser definidos, na língua portuguesa, como toda palavra que caracteriza um substantivo, indicando-lhes suas qualidades, defeitos, aspectos, estados ou aparências. Por outro lado, quando se realiza uma análise sintática, os adjetivos podem assumir diferentes funções. Assim, a palavra “trabalhador”, que pode ser caracterizada, morfológicamente, como um adjetivo na expressão “o homem trabalhador”, à medida em que caracteriza o substantivo homem, pode, sintaticamente, ser classificada como um sujeito, como em “o trabalhador chegou à empresa pela manhã”.

A eliminação dos perfis não consumidores talvez seja a principal alteração apresentada no modelo proposto e uma das principais contribuições desse trabalho. A partir de atributos dos perfis coletados e outros atributos derivados, foi possível criar um modelo de classificação de perfis, obtendo uma taxa de acerto de 86,5%. Com a aplicação desse modelo, os perfis relacionados as quatro principais montadoras foram classificados como não consumidores, além de outros perfis com conteúdo jornalístico, humorístico, referentes a personagens fictícios ou associados a outras marcas. Essa filtragem permitiu que os elementos sentimento percebido, qualidade percebida e associações à marca fossem analisados a partir das publicações de perfis classificados como consumidores, minimizando, assim, a influência de perfis

relacionados às próprias marcas estudadas ou outros perfis não consumidores, mas que eram influentes na rede.

## 5 CONSIDERAÇÕES FINAIS

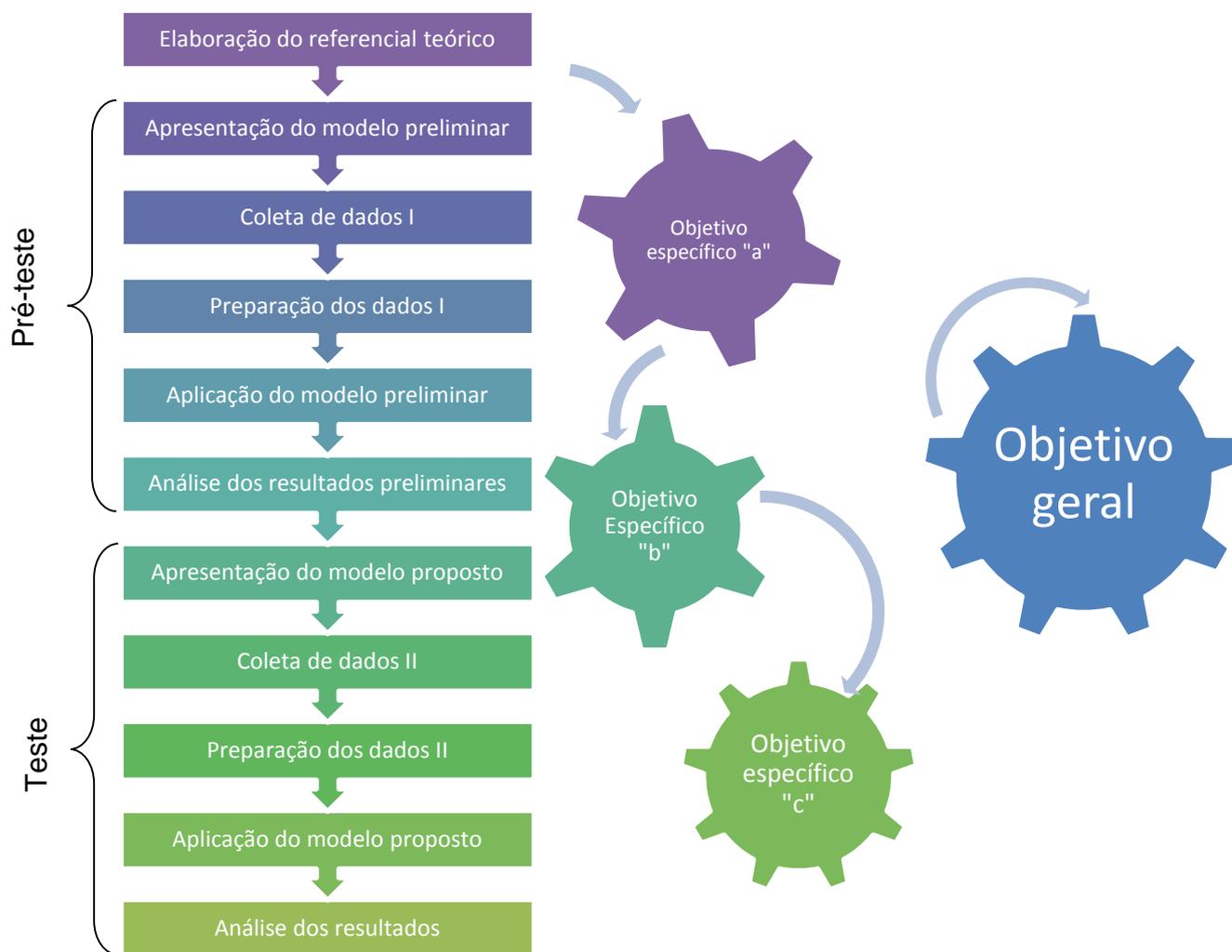
O Capítulo 5 apresenta as considerações finais referentes à pesquisa, em que são apresentadas as confrontações entre os objetivos e os resultados alcançados, as contribuições do trabalho, as limitações e sugestões para trabalhos futuros.

### 5.1 CONFRONTAÇÃO ENTRE OS OBJETIVOS E OS RESULTADOS

O objetivo geral da pesquisa consistiu em **analisar o *brand equity* na perspectiva do consumidor a partir das mídias sociais**. Para que esse objetivo fosse alcançado, três objetivos específicos foram propostos: a) identificar as dimensões presentes nos modelos conceituais de *brand equity* sob a perspectiva do consumidor; b) propor um modelo para análise do *brand equity* sob a perspectiva do consumidor a partir das mídias sociais, contemplando as dimensões identificadas nos modelos conceituais; c) validar o modelo proposto por meio de sua aplicação sobre dados gerados por usuários da rede social Twitter.

A Figura 88 apresenta as etapas executadas durante o desenvolvimento dessa pesquisa, relacionando-as com os objetivos alcançados durante esse processo.

FIGURA 88 – RELACIONAMENTO ENTRE AS ETAPAS DA PESQUISA, PASSOS EXECUTADOS, OBJETIVOS ESPECÍFICOS E OBJETIVO GERAL



FONTE: O AUTOR (2015)

O primeiro objetivo específico, **identificar as dimensões presentes nos modelos conceituais de brand equity sob a perspectiva do consumidor**, foi alcançado ainda durante a elaboração do referencial teórico, etapa em que as dimensões propostas por Keller (1993), Blackston (1992), Sharp (1996), Aaker (1998), Berry (2000) e Burmann, Jost-Benz e Riley (2009) foram apresentadas, sendo destacadas as dimensões relacionados ao conhecimento da marca, lealdade à marca, qualidade percebida e associações à marca.

O segundo objetivo específico, **propor um modelo para análise do brand equity sob a perspectiva do consumidor a partir das mídias sociais, contemplando as dimensões identificadas nos modelos conceituais**, foi atingido por meio da apresentação do modelo apresentado na Seção 4.4. Esse modelo – uma

evolução do modelo preliminar apresentado como resultado da etapa de pré-teste – buscou analisar a equidade das marcas nas redes sociais por meio de cinco elementos: conhecimento da marca, lealdade à marca, sentimento percebido, associações à marca e qualidade percebida.

Com base no modelo proposto, o conhecimento da marca normalizado pode ser obtido a partir da relação entre o grau de entrada do nó representante da marca em uma rede de seguidores e número total de nós dessa rede, de modo que quanto maior o número de seguidores, maior o conhecimento da marca na rede. O modelo apresentado propõe ainda que os valores contínuos obtidos para a dimensão conhecimento normalizado sejam transformados em valores nominais, correspondentes aos níveis da pirâmide do conhecimento proposta por Aaker(1998), de modo que os resultados possam ser mais bem compreendidos pelos profissionais de marketing.

A lealdade à marca, por sua vez, pode ser obtida por meio da relação entre o número de seguidores exclusivos da marca – aqueles que seguem o perfil da marca mas não segue nenhuma de suas concorrentes – e o número total de seguidores da marca – considerando inclusive aqueles que seguem suas concorrentes. Assim como para o critério conhecimento da marca, os valores obtidos para lealdade podem ser transformados nos níveis da pirâmide de lealdade apresentados no modelo conceitual de Aaker (1998).

O sentimento percebido pode ser dado pela relação entre o número de publicações classificadas como positivas e o número total de publicações (positivas e negativas) sobre determinada marca. Essa dimensão pode ainda ser ponderada, de acordo com o modelo proposto, pelo prestígio dos nós que compõem a rede de troca de mensagens, de forma as publicações realizadas por nós de maior prestígio na rede – calculado a partir da métrica PageRank – tem maior impacto sobre o cálculo do sentimento percebido.

As associações à marca, em seu turno, descrevem quais as palavras mais relacionadas às marcas por meio de uma nuvem de palavras, de forma que quanto maior o nível de associação das palavras às marcas, maior o tamanho dessas palavras na nuvem. A análise dessa dimensão também faz uso da classificação de polaridade das publicações, permitindo ainda a identificação das palavras mais frequentes nas publicações classificadas como positivas e naquelas classificadas como negativas.

A qualidade percebida, ao restringir o escopo morfológico das palavras mais frequentes nas publicações relacionadas às marcas, busca capturar os principais adjetivos associados às marcas nas mídias sociais, fornecendo, desse modo, subsídios para analisar a forma como a qualidade dos produtos ou serviços relacionados às marcas pesquisadas é percebida pelos consumidores presentes nessas redes.

O terceiro objetivo específico, **validar o modelo proposto por meio de aplicação sobre dados gerados por usuários da rede social Twitter**, foi alcançado por meio da execução de quatro etapas: aplicação do modelo preliminar sobre os dados coletados do Twitter referentes a quatro marcas de cosméticos presentes no Brasil: Avon, O Boticário, Jequiti e Natura; análise dos resultados obtidos pela aplicação do modelo preliminar; apresentação de um novo modelo, com base nas oportunidades de melhorias encontradas com a análise dos resultados do modelo preliminar; aplicação do modelo proposto sobre os dados coletados do Twitter referentes a quatro montadoras atuantes do Brasil: Chevrolet, Fiat, Renault e Volkswagen; por fim, análise dos resultados obtidos com a aplicação do modelo proposto, destacando suas melhorias em relação ao modelo preliminar.

Dentre as indústrias automobilísticas analisadas pelo modelo proposto, a marca Fiat atingiu o maior nível de conhecimento (0,42), superando suas concorrentes Renault (0,38), Chevrolet (0,26) e Volkswagen (0,19). Ao converter os valores obtidos para os níveis de conhecimento da pirâmide de Aaker(1998), Fiat e Renault foram categorizadas como *Top of Mind*, enquanto que Chevrolet foi tida como uma marca lembrada pelo consumidor e Volkswagen apenas como uma marca reconhecida pelos perfis presentes na rede analisada.

Apesar de ser a última colocada quando a dimensão conhecimento da marca foi analisada, a marca Volkswagen superou suas concorrentes no critério lealdade a marca (0,76), sendo seguida de perto pela Fiat (0,72), Renault (0,63) e Chevrolet (0,44). Esses resultados demonstram que os seguidores da montadora de origem alemã, apesar de poucos se comparados às suas concorrentes, são os mais leais à marca. Observou-se, entretanto, que quando os valores foram convertidos para os níveis de lealdade, as três primeiras marcas – Volkswagen, Fiat e Renault – foram categorizadas no nível em que o consumidor gosta da marca e a considera amiga, ao passo que a marca Chevrolet enquadrou-se na categoria em que o consumidor está satisfeito com a marca e é pouco vulnerável a mudanças.

Na aplicação do critério sentimento percebido houve pouca variação nas formas normalizada e ponderada. Em ambas, a marca Volkswagen foi aquela que atingiu melhores índices (próximos a 0,79) sendo seguida de perto pela Chevrolet (0,73) e Renault (0,67). A marca Fiat foi a última colocada nessa dimensão (0,46).

Dentre as associações relacionadas às marcas na rede pesquisada, destacam-se a palavra “Hall” associada à Chevrolet, indicando um retorno positivo de uma técnica de marketing denominada *naming rights* ou direitos de nome; e associações da marca Volkswagen às palavras “robô” e “fábrica”, construídas a partir da repercussão na rede de uma notícia relacionada a um acidente na fábrica da montadora na Alemanha. Destacam-se ainda: associações predominantemente negativas relacionadas a marca Fiat desencadeadas em virtude da operação lava jato em que a polícia federal apreendeu carros do ex-presidente Collor; associações predominantemente positivas da marca Renault ao modelo “Duster”, resultantes principalmente da aceitação positiva da propaganda da marca para divulgação do seu novo modelo.

Na análise da qualidade percebida, merecem destaques os adjetivos “líder”, “moderno” e “estiloso” relacionados à marca Fiat; “impressionantes” ligados à Renault; e “populares”, atribuídos pelos consumidores à marca Volkswagen.

Assim, por meio dos resultados alcançados pelo cumprimento dos três objetivos específicos supracitados, acredita-se que a pesquisa foi capaz de atingir o objetivo geral proposto, que consistia em analisar o *brand equity* na perspectiva do consumidor a partir das mídias sociais.

## 5.2 CONTRIBUIÇÕES DO TRABALHO

A principal contribuição do trabalho consiste na apresentação de um modelo para análise da equidade das marcas nas mídias sociais. Esse modelo possibilita aos profissionais de marketing e à comunidade acadêmica um melhor entendimento de como as marcas são percebidas pelos consumidores, utilizando para isso a própria estrutura da rede construída a partir das interações entre seus utilizadores, de modo completamente não incitado, ao contrário do que acontece com os modelos operacionais tradicionais, que empregam questionários para o levantamento de dados, cuja aplicação pode ser inconveniente e dispendiosa. A aplicação do modelo apresentado como resultado dessa pesquisa viabiliza ainda que diferentes marcas em

diferentes contextos sejam analisadas a partir de critérios objetivos e replicáveis que contemplam cinco dimensões da equidade da marca: conhecimento da marca, lealdade à marca, sentimento percebido, associações à marca e qualidade percebida.

Uma dessas dimensões, o sentimento percebido, trata-se de um elemento não contemplado pelos modelos tradicionais de equidade de marca. Por meio desse novo elemento, busca-se analisar a opinião geral dos consumidores em relação às marcas pesquisadas, a partir da classificação das publicações quanto a polaridade de sentimentos, que pode ser predominante positiva ou predominantemente negativa.

Além disso, a discretização dos resultados referentes ao conhecimento da marca e lealdade percebida permite uma melhor compreensão desses elementos por parte dos profissionais de marketing, à medida em que os traduz para conceitos familiares à área de atuação desses profissionais, ao contrário do modelos operacionais tradicionais, que geralmente apresentam os resultados a partir do tratamento de dados resultantes da aplicação de escalas psicométricas que buscam medir o grau de concordância de cada respondente aos itens da pesquisa.

O processo de classificação dos perfis é outra importante contribuição desse trabalho. Após a identificação, durante análise dos resultados da aplicação do modelo preliminar, de que o conjunto de usuários classificados como não-consumidores – formado principalmente por perfis de cunho jornalístico, humorístico, representantes das marcas, relacionados a personagens fictícios ou *spammers* – influenciavam diretamente nos resultados preliminares apresentados, percebeu-se que seria importante eliminar tais perfis, uma vez que a equidade da marca deveria ser analisada a partir da opinião emitida apenas por potenciais consumidores. Para a viabilização dessa tarefa, utilizou-se o algoritmo C4.5, em sua implementação J48, alcançando-se uma taxa de acerto de 86,5% no processo de classificação dos perfis.

Uma contribuição secundária do trabalho foi a realização da análise de sentimento em textos em língua portuguesa. Dado que na literatura ainda são limitadas as iniciativas nesse sentido, o roteiro apresentado nesse trabalho pode ser utilizado por outros pesquisadores que necessitem classificar textos escritos nesse idioma, já que a técnica utilizada, baseada em aprendizado de máquina, é independente de contexto.

### 5.3 LIMITAÇÕES DA PESQUISA E SUGESTÕES DE TRABALHOS FUTUROS

Essa pesquisa possui algumas limitações resultantes principalmente das técnicas empregadas para a análise dos dados, sobretudo relacionadas aos processos de classificação de sentimento, classificação dos perfis e seleção dos adjetivos relacionados às marcas.

Embora a mineração de opinião possa ser muitas vezes confundida com a atribuição de polaridade aos textos, existem outras duas tarefas relacionadas a essa técnica: mineração de opinião baseada em características e mineração de relacionamento e sentenças comparativas (PANG e LEE, 2008; LIU, 2010). A tarefa empregada nesse trabalho foi a de atribuição de polaridade, que tem como objetivo realizar a análise automática de textos avaliativos, identificando o sentimento predominante (positivo ou negativo) presente nos textos. Portanto, a técnica empregada busca identificar a opinião geral dos textos, sem analisar especificamente o sentimento atribuído a cada característica do objeto central e nem realizar a análise comparativa entre dois objetos presentes na sentença. Além disso, o processo de classificação de sentimento carrega consigo uma taxa de erro, conforme descrito na seção 3.4.2.2.4, de modo que existe a possibilidade de que o sentimento atribuído pela técnica esteja incorreto.

Buscando a redução dessa taxa de erro, sugere-se testar outras abordagens de classificação de sentimentos, como léxicas, estatísticas e semânticas, ou ainda a utilização da própria abordagem baseada em aprendizado de máquina, mas com o emprego de um conjunto de treinamento mais amplo do que o utilizado nesta pesquisa.

Assim como o processo de atribuição de polaridade, a técnica utilizada para classificação dos perfis em consumidores e não-consumidores também carrega uma taxa de erro intrínseca do próprio processo de classificação, conforme descrito na seção 3.4.2.2.6, de modo que existe a possibilidade de que a classe atribuída ao perfil pela técnica esteja incorreta. Essa classificação pode ser otimizada com a utilização de um dicionário de nomes mais completo ou ainda com a coleta de outros atributos referentes aos perfis que auxiliem na construção da árvore de classificação, minimizando a taxa de erro dessa técnica.

A seleção dos adjetivos relacionados às marcas foi realizada utilizando a análise morfológica das sentenças, ou seja, a partir da classificação das palavras em

relação a classe gramatical, sem considerar a função sintática que as palavras desempenham dentro de cada oração. Sugere-se, portanto, a adoção da análise sintática como alternativa à morfológica, de modo que a função das palavras possa estar relacionada ao contexto no qual as mesmas se encontram.

Por fim, sugere-se aplicar modelo proposto sobre dados coletados de outras mídias sociais, como Facebook e LinkedIn, por exemplo, para comparar o comportamento dos consumidores nesses diferentes ambientes sociais. O modelo proposto, que utiliza como insumo dados coletados das mídias sociais, pode ainda ser aplicado em paralelo a um ou mais modelos operacionais tradicionais sobre um mesmo conjunto de marcas, de modo que os resultados desses modelos possam ser comparados sob o mesmo prisma.

## REFERÊNCIAS

- AAKER, David A. **Criando e administrando marcas de sucesso**. São Paulo: Futura, 1996.
- \_\_\_\_\_. **Marcas: brand equity** gerenciando o valor da marca. Gulf Professional Publishing, 1998.
- \_\_\_\_\_. **Building strong brands**. Simon and Schuster, 2012.
- ACCIOLY, Anna. **Marcas de valor no mercado brasileiro**. São Paulo: Senac, 2000.
- AGRAWAL, R.; SRIKANT, Ramakrishnan. Fast Algorithms for Mining Association Rules in Large Data Bases. In: **20TH INTER-NATIONAL CONFERENCE ON VERY LARGE DATABASES**. 1994.
- ALCANTARA, Frank Coelho de. **Recuperação e classificação de informações provenientes da web e de redes sociais**. 2013. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação). Universidade Federal do Paraná. Curitiba, 2013.
- ALEXANDERSON, Gerald. About the cover: Euler and Königsberg's Bridges: A historical view. **Bulletin of the american mathematical society**, v. 43, n. 4, p. 567-573, 2006.
- ALLTON, Mike. **Social Media Active Users By Network**. Disponível em <<http://www.thesocialmediahat.com/active-users>>. Acesso em 28 abr. 2015.
- AMA, American Marketing Association. **Dictionary**. Disponível em <<https://www.ama.org/resources/Pages/Dictionary.aspx?dLetter=B>>. Acesso em 23 fev. 2015.
- AMBLER, Tim; STYLES, Chris. Brand development versus new product development: toward a process model of extension decisions. **Journal of Product & Brand Management**, v. 6, n. 4, p. 222-234, 1997.
- ANDREOLA, Rafaela. **Support vector machines na classificação de imagens hiperespectrais**. Dissertação (Mestrado em Sensoriamento Remoto). Universidade Federal do Rio Grande do Sul. Porto Alegre, 2009.
- ANGELONI, Maria Terezinha. Elementos intervenientes na tomada de decisão. **Ci. Inf**, v. 32, n. 1, p. 17-22, 2003.
- ARMSTRONG, Gary; KOTLER, Philip. **Princípios de marketing**. São Paulo: Prentice Hall, 2004.
- BARNES, John Arundel. **Class and committees in a Norwegian island parish**. Plenum, 1954.
- BARNES, John Arundel; HARARY, Frank. Graph theory in network analysis. **Social Networks**, v. 5, n. 2, p. 235-244, 1983.

BARREDA, Albert. Creating brand equity when using travel-related online social network Web sites. **Journal of Vacation Marketing**, p. 1356766714532465, 2014.

BECKER, Karin; TUMITAN, Diego. Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 28., 2013, Recife. **Anais...** Recife: SBC, 2013.

BENEVENUTO, Fabrício; ALMEIDA, Jussara M.; SILVA, Altigran S. Explorando redes sociais online: Da coleta e análise de grandes bases de dados às aplicações. In: MINI-CURSOS DO SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS, 23., 2011, Vitória. **Anais...** Vitória: SBC, 2011.

BERMINGHAM, Adam; SMEATON, Alan F. Classifying sentiment in microblogs: is brevity an advantage?. In: **PROCEEDINGS OF THE 19TH ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT**. ACM, 2010. p. 1833-1836.

BERRY, Leonard L. Cultivating service brand equity. **Journal of the Academy of marketing Science**, v. 28, n. 1, p. 128-137, 2000.

BHUIYAN, Touhid; XU, Yue; JOSANG, Audun. State-of-the-art review on opinion mining from online customers' feedback. In: **PROCEEDINGS OF THE 9TH ASIA-PACIFIC COMPLEX SYSTEMS CONFERENCE**. Chuo University, 2009. p. 385-390.

BLACKSTON, Max. Observations: building brand equity by managing the brand's relationships. **Journal of advertising research**, v. 32, n. 3, p. 79-83, 1992.

BOIY, Erik; MOENS, Marie-Francine. A machine learning approach to sentiment analysis in multilingual Web texts. **Information retrieval**, v. 12, n. 5, p. 526-558, 2009.

BOLLEN, Johan; MAO, Huina; ZENG, Xiaojun. Twitter mood predicts the stock market. **Journal of Computational Science**, v. 2, n. 1, p. 1-8, 2011.

BONDY, John Adrian; MURTY, Uppaluri Siva Ramachandra. **Graph theory with applications**. London: Macmillan, 1976.

BORGATTI, Stephen P. **Basic Social Network Concepts**. 2002. Disponível em: <<http://www.analytictech.com/borgatti/.com/ucinet.htm>>. Acesso em: 10 mar. 2015.

BOYD, Danah; GOLDBER, Scott; LOTAN, Gilad. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: **SYSTEM SCIENCES (HICSS), 2010 43RD HAWAII INTERNATIONAL CONFERENCE ON**. IEEE, 2010. p. 1-10.

BURGET, Radim et al. Rapidminer image processing extension: A platform for collaborative research. In: **PROCEEDINGS OF THE 33RD INTERNATIONAL CONFERENCE ON TELECOMMUNICATION AND SIGNAL PROCESSING**. 2010. p. 114-118.

BURMANN, Christoph; JOST-BENZ, Marc; RILEY, Nicola. Towards an identity-based brand equity model. **Journal of Business Research**, v. 62, n. 3, p. 390-397, 2009.

CAMBRIA, Erik et. al. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, v. 28, n. 2, p. 15-21, 2013.

CAMPOS, Ronaldo Ribeiro de. **Redes Complexas e ações para compartilhamento de conhecimento**: uma análise de redes sociais em um ambiente Web para apoio à aprendizagem. Tese (Doutorado em Engenharia de Produção). Universidade de São Paulo. São Carlos, 2014.

CASTRO, Izabel Cristina Andion. **Uma ferramenta para análise da comunicação organizacional através das redes sociais**. 2005. Dissertação (Mestrado Profissional em Redes de Computadores). Universidade Salvador. Salvador, 2005.

CHANDOLA, Varun; BANERJEE, Arindam; KUMAR, Vipin. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, v. 41, n. 3, p. 15, 2009.

CHAVES, Eduardo de Paula e Silva. **Brand equity na perspectiva dos clientes**: uma investigação empírica de suas dimensões e do mix de marketing entre usuários de marca de sapato esportivo. 2010. Dissertação (Mestrado em Administração de Organizações). Universidade de São Paulo, 2010.

CHEN, Hsinchun; ZIMBRA, David. AI and opinion mining. *Intelligent Systems, IEEE*, v. 25, n. 3, p. 74-80, 2010.

CHRISTODOULIDES, George; DE CHERNATONY, Leslie. Consumer-based brand equity conceptualization and measurement: A literature review. *International journal of research in marketing*, v. 52, n. 1, p. 43-66, 2010.

CIOU, K. et. al. **Data Mining: A Knowledge Discovery Approach**. New York: Springer, 2007.

CIRIACO, Douglas. **O que é API?** 2009. Disponível em: <<http://www.tecmundo.com.br/programacao/1807-o-que-e-api-.htm>>. Acesso em: 10 fev. 2015.

CUNHA, Celso; CINTRA, Luís F. Lindley. **Nova gramática do português contemporâneo**. Rio de Janeiro: Nova Fronteira, 2008.

D'EMÍDIO, Marcelo. **O valor da marca**: uma abordagem de equações estruturais. Tese (Doutorado em Administração). Universidade de São Paulo. São Paulo. 2009.

DAVENPORT, Thomas H. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. Futura, 2000.

DAVENPORT, Thomas H.; PRUSAK, Laurence. **Conhecimento empresarial**. 3. ed. Rio de Janeiro: Campus, 1998.

DD-CSS. **About DD-CSS**. Disponível em <<http://dd-css.com/docs/>>. Acesso em: 14 set. 2015.

DE FREITAS, Mariza Branco Rodrigo. **Valor da marca**: um estudo sobre a problemática de sua evidenciação contábil no Brasil. Dissertação (Mestrado em Ciências Contábeis). Universidade do Estado do Rio de Janeiro. Rio de Janeiro. 2012.

DE NOOY, Wouter; MRVAR, Andrej; BATAGELJ, Vladimir. **Exploratory social network analysis with Pajek**. Cambridge University Press, 2011.

DIAS, Tatiana de Mello. Facebook ultrapassa audiência do Google. **O Estado de S. Paulo**. São Paulo, 17 abr. 2012. Disponível em: <<http://blogs.estadao.com.br/link/facebook-ultrapassa-audiencia-do-google/>>. Acesso em: 02 jun. 2014.

DRUCKER, Peter. **Desafios gerenciais para o século XXI**. São Paulo: Cengage Learning Editores, 2006.

ELLISON, Nicole B. et. al. Social network sites: Definition, history, and scholarship. **Journal of Computer-Mediated Communication**, v. 13, n. 1, p. 210-230, 2007.

EMARKETEER. **Number of social network users in selected countries in 2011 and 2017** (in millions). Disponível em: <<http://www.statista.com/statistics/278341/number-of-social-network-users-in-selected-countries/>>. Acessado em: 02 jun. 2014.

EPOCA. **Robô mata funcionário da Volkswagen na Alemanha**. Disponível em <http://epoca.globo.com/vida/experiencias-digitais/noticia/2015/07/robo-mata-funcionario-da-volkswagen-na-alemanh.html>>. Acessado em: 5 out. 2015.

ESULI, Andrea; SEBASTIANI, Fabrizio. Determining the semantic orientation of terms through gloss classification. In: **PROCEEDINGS OF THE 14TH ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT**. ACM, 2005. p. 617-624.

EULER, Leonhard. **Solutio Problematis ad Geometriam Situs Pertinentis**. Commentarii Academiae Scientiarum Imperialis Petropolitanae 8, 128-140, 1736.

FARQUHAR, Peter H. Managing brand equity. **Journal of Advertising Research**, v. 30, n. 4, p. 7-12, 1989.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996.

FRANÇA, Andre Luiz Dias de. **A estrutura do fluxo informacional do sistema nacional de transplantes: uma investigação sob a óptica da análise de redes sociais**. Dissertação (Mestrado em Ciência da Informação). Universidade Federal da Paraíba. João Pessoa, 2012.

FRAWLEY, William J.; PIATETSKY-SHAPIRO, Gregory; MATHEUS, Christopher J. Knowledge discovery in databases: An overview. **AI magazine**, v. 13, n. 3, p. 57, 1992.

FREEMAN, Linton C. Some antecedents of social network analysis. **Connections**, v. 19, n. 1, p. 39-42, 1996.

FREITAS, Marcos Cezar; PEREIRA, Hernane Borges de Barros. Contribuição da análise de redes sociais para o estudo sobre os fluxos de informações e

conhecimento. In: ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO: INFORMAÇÃO, CONHECIMENTO E SOCIEDADE DIGITAL, 6., Salvador, 2005. **Anais...** Salvador: UFBA, 2005.

GIL, Antonio Carlos. Métodos e técnicas de pesquisa social. In: **Métodos e técnicas de pesquisa social**. Atlas, 2010.

GO, Alec; BHAYANI, Richa; HUANG, Lei. Twitter sentiment classification using distant supervision. **CS224N Project Report**, Stanford, p. 1-12, 2009.

GOLUB, Gene H.; VAN LOAN, Charles F. **Matrix Computations**. JHU Press, 2012.

GRACZYK, Magdalena; LASOTA, Tadeusz; TRAWIŃSKI, Bogdan. Comparative analysis of premises valuation models using KEEL, RapidMiner, and WEKA. In: **COMPUTATIONAL COLLECTIVE INTELLIGENCE. SEMANTIC WEB, SOCIAL NETWORKS AND MULTIAGENT SYSTEMS**. Springer Berlin Heidelberg, 2009. p. 800-812.

GRANNELL, Chris. Untangling Brand Equity, Value, Assets and Healt. **Brand Papers**, 2009.

GRIBKOVSKAIA, Irina et al. General solutions to the single vehicle routing problem with pickups and deliveries. **European Journal of Operational Research**, v. 180, n. 2, p. 568-584, 2007.

HAIGH, David. Understanding the financial value of brands. **European Association of Advertising Agencies Raporu**, 1999.

HALL, Mark et al. The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**, v. 11, n. 1, p. 10-18, 2009.

HAN, J. Kamber; KAMBER, Micheeline. M. **Data mining: concepts and techniques**. University of Illinois at Urbana-Champaign, 2006.

HANNEMAN, Robert A.; RIDDLE, Mark. **Introduction to social network methods**. Riverside: University of California, 2005.

HANSEN, Derek; SHNEIDERMAN, Ben; SMITH, Marc A. **Analyzing social media networks with NodeXL: Insights from a connected world**. Morgan Kaufmann, 2010.

HARARY, Frank. **Graph theory**. Reading: Addison-Wesley, 1969.

HARTIGAN, John A.; WONG, Manchek A. Algorithm AS 136: A k-means clustering algorithm. **Applied statistics**, p. 100-108, 1979.

HOFMANN, Markus; KLINKENBERG, Ralf (Ed.). RapidMiner: Data mining use cases and business analytics applications. **CRC Press**, 2013.

HSU, Lih-Hsing; LIN, Cheng-Kuan. Graph theory and interconnection networks. **CRC press**, 2008.

HSU, Lih-Hsing; LIN, Cheng-Kuan. **Graph theory and interconnection networks**. CRC press, 2008.

HU, Mingqing; LIU, Bing. Mining and summarizing customer reviews. In: **PROCEEDINGS OF THE TENTH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING**. Seattle: ACM, 2004. p. 168-177.

HUBERMAN, Bernardo A.; ROMERO, Daniel M.; WU, Fang. Social networks that matter: Twitter under the microscope. **Available at SSRN 1313405**, 2008.

IBGE. Instituto Brasileiro de Geografia. Estudos e Pesquisas. **Projeção da População do Brasil por Sexo e Idade – 1980-2050**. Revisão 2008. Rio de Janeiro, 2008. Disponível em: <[http://www.ibge.gov.br/home/estatistica/populacao/projecao\\_da\\_populacao/2008/projecao.pdf](http://www.ibge.gov.br/home/estatistica/populacao/projecao_da_populacao/2008/projecao.pdf)>. Acesso em: 02 jun. 2014.

JANSEN, Bernard J. et al. Twitter power: Tweets as electronic word of mouth. **Journal of the American society for information science and technology**, v. 60, n. 11, p. 2169-2188, 2009.

JAVA, Akshay et al. Why we twitter: understanding microblogging usage and communities. In: **PROCEEDINGS OF THE 9TH WEBKDD AND 1ST SNA-KDD 2007 WORKSHOP ON WEB MINING AND SOCIAL NETWORK ANALYSIS**. ACM, 2007. p. 56-65.

JOHN, George H.; LANGLEY, Pat. Estimating continuous distributions in Bayesian classifiers. In: **PROCEEDINGS OF THE ELEVENTH CONFERENCE ON UNCERTAINTY IN ARTIFICIAL INTELLIGENCE**. Morgan Kaufmann Publishers Inc., 1995. p. 338-345.

KAMAKURA, Wagner A.; RUSSELL, Gary J. Measuring brand value with scanner data. **International Journal of Research in Marketing**, v. 10, n. 1, p. 9-22, 1993.

KAMPS, Jaap et al. Using WordNet to Measure Semantic Orientations of Adjectives. In: **LREC**. 2004. p. 1115-1118

KAPFERER, Jean-Nöel. **The new strategic brand management: Advanced insights and strategic thinking**. Londres: Kogan Page Publishers, 2012.

KAVEH, Ali. **Optimal analysis of structures by concepts of symmetry and regularity**. New York: Springer, 2013.

KELLER, Kevin Lane. Conceptualizing, measuring, and managing customer-based brand equity. **The Journal of Marketing**, p. 1-22, 1993.

KELLER, Kevin Lane; MACHADO, Marco. **Gestão estratégica de marcas**. 2006.

KIM, P.; ANDERSON, E.; JOSEPH, J. The forrester wave: Brand monitoring. **Cambridge: Forrester Research**, 2006.

KIM, Soo-Min; HOVY, Eduard. Determining the sentiment of opinions. In: Proceedings of the 20th international conference on Computational Linguistics. **Association for Computational Linguistics**, 2004. p. 1367.

KOTLER, Philip. **Marketing para o século XXI: como criar, conquistar e dominar mercados**. São Paulo: Futura, 2000.

KOTLER, Philip; KELLER, Kevin Lane. **Administração de marketing**. São Paulo: Prentice Hall, 2006.

KWAK, Haewoon et al. What is Twitter, a social network or a news media?. In: **PROCEEDINGS OF THE 19TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB**. ACM, 2010. p. 591-600.

LAROSE, Daniel T. **Discovering Knowledge in Data: An Introduction to Data mining**. New Jersey: John Wiley & Sons. Inc., Publication, 2005.

LASSAR, Walfried; MITTAL, Banwari; SHARMA, Arun. Measuring customer-based brand equity. **Journal of consumer marketing**, v. 12, n. 4, p. 11-19, 1995.

LEUTHESSER, Lance. Defining, Measuring, and Managing Brand Equity: A Conference Summary by. **Marketing Science Institute**, 1988.

LÉZORAY, Olivier; GRADY, Leo (Ed.). **Image processing and analysis with graphs: theory and practice**. Boca Raton: CRC Press, 2012.

LIMA, Ana Carolina Espírito Santo. **Análise de Sentimento e Desambiguação no Contexto da TV Social**. 2012. Dissertação (Mestrado em Engenharia Elétrica). Universidade Presbiteriana Mackenzie. São Paulo, 2012.

LIMA, Gisele Passos. **A percepção dos profissionais de assessorias de comunicação frente aos sites de redes sociais**. 2011. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação). Universidade Federal do Paraná. Curitiba, 2011.

LIU, Bing. **Web data mining: exploring hyperlinks, contents, and usage data**. Springer Science & Business Media, 2007.

\_\_\_\_\_. Sentiment analysis and subjectivity. **Handbook of Natural Language Processing**, v. 2, p. 627-666, 2010.

\_\_\_\_\_. Sentiment analysis and opinion mining. **Synthesis Lectures on Human Language Technologies**, v. 5, n. 1, p. 1-167, 2012.

LIU, Bing; HSU, Wynne; MA, Yiming. Integrating Classification and Association Rule Mining. In: **FOURTH INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING**, 80-86, 1998.

LIU, Bing; YU, Philip S. Page Rank. The Top 10 Algorithms in Data Mining. **Knowledge and Information Systems**, v. 14, n. 1, p. 1-37, 2008.

LIU, Bing; ZHANG, Lei. A survey of opinion mining and sentiment analysis. In: **Mining text data**. New York: Springer, 2012. p. 415-463

LOBUR, M. et. al. Some trends in knowledge discovery and data mining. In: **PERSPECTIVE TECHNOLOGIES AND METHODS IN MEMS DESIGN, 2008. MEMSTECH 2008.INTERNATIONAL CONFERENCE ON**. IEEE, 2008. p. 95-97.

MALHOTRA, Yogesh. **Role of information technology in managing organizational change and organizational interdependence**. BRINT Institute, LLC, New York, 1993.

MARTINS, José Roberto. **Branding: um manual para você criar, avaliar e gerenciar marcas**. São Paulo: Negócio Editora, 2000.

MARWICK, Alice E. et al. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. **New media & society**, v. 13, n. 1, p. 114-133, 2011.

MILLER, George A. et. al. Introduction to wordnet: An on-line lexical database\*. **International journal of lexicography**, v. 3, n. 4, p. 235-244, 1990.

MIRANDA FILHO, Renato. **Um arcabouço para pesquisas de opinião em redes sociais**. Universidade Federal de Minas Gerais. 2014. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Minas Gerais. Belo Horizonte. 2014.

MITCHELL, Tom M. **Machine learning**. 1997. Burr Ridge, IL: McGraw Hill, v. 45, 1997

MORENO, Jacob Levy. **Who shall survive?: A new approach to the problem of human interrelations**. University of Michigan: Nervous and mental disease publishing co, 1934.

MÜLLER-PROTHMANN, Tobias. **Social network analysis: A practical method to improve knowledge sharing**. Hands-on knowledge co-creation and sharing, p. 219-233, 2007.

NAM, Hyoryung; KANNAN, P. K. The informational value of social tagging networks. **Journal of Marketing**, v. 78, n. 4, p. 21-40, 2014.

NODEXL: Network Overview, Discovery and Exploration for Excel. Disponível em:<<http://nodexl.codeplex.com>>. Acesso em: 19 ago. 2015.

NOOY, Wouter; MRVAR, Andrej; BATAGELJ, Vladimir. **Exploratory Social Network Analysis with Pajek**. Cambridge: Cambridge University Press, 2005.

OGURI, P. **Aprendizado de máquina para o problema de sentiment classification**. Dissertação (Mestrado em Informática). Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, 2006.

OLIVEIRA, Marta Olívia Rovedder de. **Modelo de estimação de Brand Equity**. 2009. Tese (Doutorado em Administração). Universidade Federal do Rio Grande do Sul. Porto Alegre. 2013.

OLIVEIRA, Saulo Campos. **Redes de colaboração científica**: a dinâmica da rede em nanotecnologia. Dissertação (Mestrado em Ciência, tecnologia e sociedade). Universidade Federal de São Carlos. São Carlos, 2011.

OLIVER, Richard L. **Satisfaction**: A behavioral perspective on the consumer. Routledge, 2014.

OLSON, David L.; DELEN, Dursun. **Advanced data mining techniques**. Springer Science & Business Media, 2008.

OTTEY, Matthew Eric; PARTHASARATHY, Srinivasan; GHOTING, Amol. An empirical comparison of outlier detection algorithms. **Data Mining Methods for Anomaly Detection**, p. 45, 2005.

PAGE, Lawrence et. al. **The PageRank citation ranking**: bringing order to the Web. Stanford: Stanford InfoLab Publication Server, 1999.

PAK, Alexander; PAROUBEK, Patrick. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: **LREC**. 2010. p. 1320-1326.

PANG, Bo; LEE, Lillian. Opinion mining and sentiment analysis. **Foundations and trends in information retrieval**, v. 2, n. 1-2, p. 1-135, 2008.

PANG, Bo; LEE, Lillian; VAITHYANATHAN, Shivakumar. Thumbs up?: sentiment classification using machine learning techniques. In: **PROCEEDINGS OF THE ACL-02 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING**. Stroudsburg: Association for Computational Linguistics, 2002. p. 79-86.

PAPPU, Ravi; QUESTER, Pascale G.; COOKSEY, Ray W. Consumer-based brand equity: improving the measurement-empirical evidence. **Journal of Product & Brand Management**, v. 14, n. 3, p. 143-154, 2005.

PARK, Chan Su; SRINIVASAN, Veena. A survey-based method for measuring and understanding brand equity and its extendibility. **Journal of marketing research**, p. 271-288, 1994.

PIATETSKY-SHAPIO, Gregory. Knowledge discovery in real databases: A report on the IJCAI-89 Workshop. **AI magazine**, v. 11, n. 4, p. 68, 1990.

PIRES, Diogo Manoel Simões. **O impacto da promoção de preços sobre o consumer-based brand equity**: um estudo experimental a partir de anúncios impressos. 2013. Dissertação (Mestrado em Administração). Universidade Federal do Rio Grande do Sul. Porto Alegre. 2013.

QUINLAN, J. Ross . Induction of decision trees. **Machine learning**, v. 1, n. 1, p. 81-106, 1986.

\_\_\_\_\_. **C4. 5**: programs for machine learning. San Francisco: Morgan Kaufmann Publishers, 1993.

RAGGIO, Randle D.; LEONE, Robert P. The theoretical separation of brand equity and brand value: Managerial implications for strategic planning. **Journal of Brand Management**, v. 14, n. 5, p. 380-395, 2007.

RAMAMOHAN, Y. et al. A Study of Data Mining Tools in Knowledge Discovery Process. **International Journal of Soft Computing and Engineering (IJSCE) ISSN**, p. 2231-2307, 2012.

REUTENAUER, Coralie; URVOY, Tanguy. Plateforme d'étiquetage dynamique pour la fouille d'opinion. **Document Numérique**, v. 17, n. 1, p. 85-101, 2014.

RICHTER, Felix. **How marketers use social media**. 2014a. Disponível em <<http://www.statista.com/chart/2289/how-marketers-use-social-media/>>. Acesso em 28 abr. 2015.

\_\_\_\_\_. **Brands use Twitter to boost awareness rather than sales**. 2014b. Disponível em <<http://www.statista.com/chart/2058/reasons-to-use-twitter-as-a-marketing-tool/>>. Acesso em 28 abr. 2015.

RITZMANN, Bárbara Nascimento Barbosa. **Redes sociais online como fontes de informação**. 2012. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação). Universidade Federal do Paraná. Curitiba, 2012.

RUOHONEN, Keijo. **Graph Theory**. Finlândia: Tampere University of Technology, 2013.

RUST, Roland T.; LEMON, Katherine N.; ZEITHAML, Valarie A. Return on marketing: using customer equity to focus marketing strategy. **Journal of marketing**, v. 68, n. 1, p. 109-127, 2004.

SANTOS, Denise Cristiane dos. **Coleta automatizada e análise de dados em fanpages do facebook**. 2014. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação). Universidade Federal do Paraná. Curitiba, 2014.

SANTOS, Fernão Reges dos. **Comunidades virtuais baseadas em vídeo digital: uma proposta de conteúdo adaptativo pautada em redes de aprendizagem e agentes inteligentes**. Dissertação (Mestrado em Engenharia Elétrica). Universidade Presbiteriana Mackenzie. São Paulo, 2010.

SANTOS, Robson; DE PRÓSPERO, Raphael Piedade; GIANORDOLI, Gabriel. Estudo sobre a abordagem quantitativa para visualização de dados qualitativos: processo e ferramentas. **Blucher Design Proceedings**, v. 1, n. 2, 2014.

SCHULTZ, Don E.; BARNES, Beth E. **Campanhas estratégicas de comunicação de marca**. Rio de Janeiro: Qualitymark, 2001.

SHANKAR, Venkatesh; AZAR, Pablo; FULLER, Matthew. Practice prize paper-BRAN\* EQT: A multicategory brand equity model and its application at allstate. **Marketing Science**, v. 27, n. 4, p. 567-584, 2008.

SHARP, Byron. Brand equity and market-based assets of professional service firms. **Journal of Professional Services Marketing**, v. 13, n. 1, p. 3-13, 1996.

SIMON, Carol J.; SULLIVAN, Mary W. The measurement and determinants of brand equity: a financial approach. **Marketing science**, v. 12, n. 1, p. 28-52, 1993.

SOARES, José Carlos Tinoco. **Tratado da propriedade industrial: patentes e seus sucedâneos**. Brasília: Jurídica Brasília, 1998.

SOUZA, Leandro Henrique de. **Redes sociais como proposta para amplificar a criação do conhecimento em organizações inovadoras**. 2010. Dissertação (Mestrado em Ciência, Gestão e Tecnologia da Informação). Universidade Federal do Paraná. Curitiba, 2010.

SRIVASTAVA, Rajendra K.; SHERVANI, Tasadduq A.; FAHEY, Liam. Market-based assets and shareholder value: a framework for analysis. **The Journal of Marketing**, p. 2-18, 1998.

SUMATHI, Sai; SIVANANDAM, S. N. **Introduction to data mining and its applications**. Springer, 2006.

SWAIT, Joffre et al. The equalization price: A measure of consumer-perceived brand equity. **International Journal of Research in Marketing**, v. 10, n. 1, p. 23-45, 1993.

SYMEONIDIS, Andreas L.; MITKAS, Pericles A. **Agent intelligence through data mining**. New York: Springer Science & Business Media, 2006.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao datamining: mineração de dados**. Ciencia Moderna, 2009.

TAVARES, Mauro Calixta. **A força da marca: como construir e manter marcas fortes**. São Paulo: Habra, 1998.

THORAT, Surabhi. **Opinion Mining and Sentiment Analysis-Its Tools and Challenges**. 2014.

TRAVERSO, Martin. **Presto: Interacting with petabytes of data at Facebook**. Menlo Park, 06 nov. 2013. Disponível em: <<https://www.facebook.com/notes/facebook-engineering/presto-interacting-with-petabytes-of-data-at-facebook/10151786197628920>>. Acessado em: 02 jun. 2014.

TSYTSARAU, Mikalai; PALPANAS, Themis. Survey on mining subjective data on the web. **Data Mining and Knowledge Discovery**, v. 24, n. 3, p. 478-514, 2012.

TUMASJAN, Andranik et. al. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. **ICWSM**, v. 10, p. 178-185, 2010.

TURNEY, Peter D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: **PROCEEDINGS OF THE 40TH ANNUAL MEETING ON ASSOCIATION FOR COMPUTATIONAL LINGUISTICS**. Association for Computational Linguistics, 2002. p. 417-424.

TWITTER, Inc. **New Tweets per second record, and how!**São Francisco, 16 ago. 2013. Disponível em: <<https://blog.twitter.com/2013/new-tweets-per-second-record-and-how>>. Acessado em 02 jun. 2014.

VAPNIK, Vladimir. **Estimation of dependences based on empirical data**. New York: Springer Science & Business Media, 2006.

VAZQUEZ, Rodolfo; DEL RIO, A. Belen; IGLESIAS, Victor. Consumer-based brand equity: development and validation of a measurement instrument. **Journal of Marketing management**, v. 18, n. 1-2, p. 27-48, 2002.

VERGARA, Sylvia Constant. **Projetos e relatórios de pesquisa em administração**. Atlas, 1997.

VILLANUEVA, Julian; HANSSENS, Dominique M. **Customer equity: Measurement, management and research opportunities**. Hanover: Now Publishers , 2007.

VINODHINI, G.; CHANDRASEKARAN, R. M. Sentiment analysis and opinion mining: a survey. **International Journal**, v. 2, n. 6, 2012.

WASSERMAN, Stanley; FAUST, Katherine. **Social network analysis: Methods and applications**. Cambridge university press, 1994.

YOO, Boonghee; DONTU, Naveen. Developing and validating a multidimensional consumer-based brand equity scale. **Journal of business research**, v. 52, n. 1, p. 1-14, 2001.

ZEITHAML, Valarie A. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. **The Journal of marketing**, p. 2-22, 1988.